

Heterologous protein production in *Streptomyces lividans*

Rattleff, Stig; Thykær, Jette; Eliasson Lantz, Anna

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Rattleff, S., Thykær, J., & Eliasson Lantz, A. (2013). Heterologous protein production in *Streptomyces lividans*. Kgs. Lyngby: Technical University of Denmark (DTU).

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

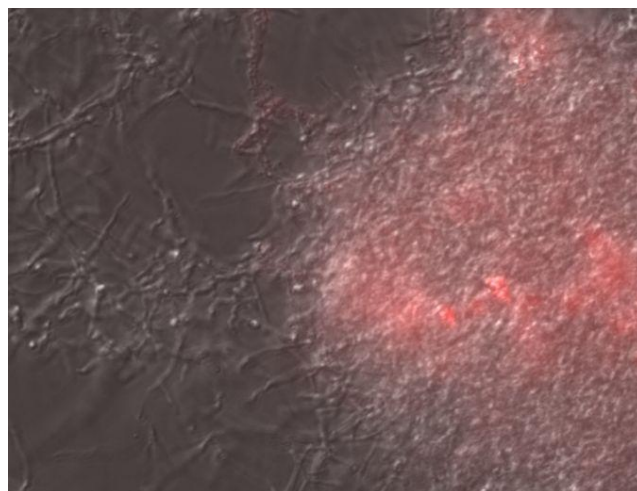
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Heterologous protein production in *Streptomyces* *lividans*

PhD thesis

Stig Rattleff

June 2013



Preface

Abstract

Alternative hosts for expressing heterologous proteins have recently gained an increased attention. *Actinomycetes*, and among these especially the *Streptomyces* are considered promising candidates, since they through their saprophytic lifestyle are capable of excreting large amounts of proteins. Industrially this is very useful due to the reduction of downstream processing. *Streptomyces* have long been studied, and a great amount of knowledge has been gained on genetic tools and metabolism. A most promising candidate as host among the *Streptomyces* is *S. lividans*, since this strain exhibits an exceptionally low protease activity, ensuring good product stability. Despite the fact that *S. lividans* has already seen industrial application studies on quantitative physiology are still lacking. It will greatly benefit the use as a common host to elucidate how *S. lividans* behaves in submerged cultivations, as well as how it is affected by expressing a foreign protein.

In this thesis methods have been established for the study of quantitative physiology and a method for screening large amounts of carbon/nitrogen/phosphorus sources have been tested. Further, parallel to the project that is the basis of the thesis is an EU project that has had to be supported with physiological data. Cultivations have been run across laboratory relevant scales. A strong physiology platform is required to obtain the most informative results from the advanced 'omics' analyses. Whether the biomass is harvested from batch or chemostat experiments, reproducibility is essential for comparing results under controlled conditions.

Dansk Resumé

Antallet af værter der med fordel kan bruges til heterolog protein produktion har været søgt forøget gennem mange år. *Actinomyceterne*, heriblandt særligt *Streptomyces* er lovende kandidater, da de gennem deres naturlige livsstil er i stand til at udskille store mængder proteiner. Dette søges udnyttet industrielt, da extracellulær secernering i høj grad minimerer mængden af oprensning, og dermed er omkostning minimerende. Som værter har *Streptomyces* været undersøgt gennem en årrække, og der er opnået stor viden indenfor genetiske værktøjer og metabolisme. En klar kandidat som vært blandt *Streptomyces* er *S. lividans* da denne har en exceptionelt lav extracellulær protease aktivitet, og hermed kan bedre produkt stabilitet sikres. På trods af at *S. lividans* har været brugt industrielt, mangler der dog stadig nogle studier om kvantitativ fysiologi, da det vil øge brugbarheden som vært at få klarlagt hvordan *S. lividans* opfører sig i kultiveringer, og at få estimeret hvordan cellen påvirkes af at udtrykke et fremmed protein.

I denne afhandling er metoder blevet etableret, og dermed er fundamentet lagt for studier af den kvantitative fysiologi. Der er yderligere blevet testet en metode til screening af anvendelige kulstof/nitrogen/fosfat kilder. Ydermere har projektet der ligger til grund for afhandlingen været kørt parallelt med et EU projekt, og dette skulle også støttes med kvantitative studier. Der er blevet udført kultiveringer på tværs af laboratorierelevante skala. En stærk forståelse af fysiologi er nødvendig for at sikre de mest sigende resultater når de avancerede 'omics' metoder benyttes. Uanset om biomassen bliver høstet fra batch eller chemostat forsøg er reproducerbarhed en forudsætning for at kunne sammenholde resultater der er opnået under kontrollerede forhold.

Acknowledgements

In a project of this magnitude there are many people to thank, and the safest bet would be to say “none mentioned, none forgotten”, however some people have deserved a special mention.

A special place is reserved for my wife, Mai, who has shown exceptional patience with me, throughout the project, and especially through the writing phase. Fermentation technology is a field of study that can be difficult to unite with having a family, but you have stood through it, and for that I am extremely grateful.

I have been blessed with a lot of good colleagues that have somehow kept their heads over the water in these turbulent times, where renovation of our building has caused a great deal of inconvenience to all of us, and later a strategic restructuring. I wish you the best.

To all of the colleagues here at CMB, you have been among the people that have made this work place so very inspiring. There are obviously many people who have finished their time at DTU before me, and while not being mentioned, you are not forgotten.

Peter, I would also like to thank you for helping me with the very last experiments.

Subir, we have shared an interesting organism, and many talks on all the opportunities, as well as difficulties this entails.

Jette T. I would like to thank you for fighting hard to keep me from despairing every time there is yet another hopeless incident of equipment failure or the like.

Anna E. I would like to thank you for the opportunity to perform my PhD here at DTU.

Martin, Tina, and Elisabeth, you together with the administrative people, Birgitte, Birgitte, and Anne, as well as past administrative people, have been the life blood of this centre. All the things that have needed fixing at hopeless hours, autoclaves not working, missing chemicals, travel expense management, working atmosphere and so forth. You have been the people that somehow have made this centre run during all of the structural changes. Thank you for being good colleagues, and invaluable life-savers.

Bibliography

Nandy, Subir Kumar, Stig Rattleff, Gernay Krist, and Anna Eliasson Lantz. “Model of enhanced heterologous protein production by *Streptomyces lividans* – Effect of C/N ratio.” *Manuscript in prep*, 2013.

Rattleff, Stig, Jens Nielsen, Stefan Rokem, and Anna Eliasson Lantz. “Metabolic flux profiling of sixteen Actinobacteria species.” *Manuscript in prep*, 2013b.

Rattleff, Stig, Jette Thykaer, and Anna Eliasson Lantz. “Rational selection of nitrogen source for optimized heterologous protein production using *Streptomyces lividans*.” *Applied Microbiology and Biotechnology*, 2013a.

Special notes to the bibliography

For the (Nandy, et al. 2013) paper, I have contributed with the experimental work, that is I have performed the fermentations, and provided the glucose and mRFP analyses. And I have had a minor contribution to the manuscript. The manuscript is still in preparation and not included.

For (Rattleff, Nielsen, Rokem, & Lantz, 2013b), I have contributed both experimentally, to the data analysis and to the manuscript. The paper is included as appendix C

(Rattleff, Thykaer, & Lantz, 2013a) is a paper I have included as a separate chapter.

A final note

The entire work included in the PhD thesis has been carried out at DTU department for Systems Biology, Centre for Microbial Biotechnology (CMB), under the supervision of Anna E. Lantz and Jette Thykær.

Contents

Preface.....	1
Abstract	1
Dansk Resumé	1
Acknowledgements	2
Bibliography	2
Special notes to the bibliography	3
A final note	3
Chapter 1. Outline	8
Works Cited	11
Chapter 2. Introduction	12
Industrial biotechnology.....	12
Product complexity.....	13
Process/organism complexity.....	14
Microbial cell factories	15
Heterologous protein production	15
Endogenous vs. heterologous host	16
Well established heterologous hosts for protein production	17
Bacterial hosts	17
Yeast hosts.....	17
Fungal hosts.....	18
Streptomyces – state of the art	19
Works Cited	22
Chapter 3. Screening and growth of streptomyces	26
Establishing a propagation procedure.....	29
Germination.....	29
Germination time	31
Growth experiments in batch.....	32
Initial characterisation	33
Using an amino acid as nitrogen source.....	34
Influence of thiostrepton concentration	35
Conclusions on the batch experiments	37
Extending the range of carbon sources, functional genomics.	37

Carbon source screening	38
Hierarchical clustering	42
Conclusions C- sources	43
Summing up	44
Materials and methods	44
Sporestock	44
Germination.....	45
Inoculation conc.	45
Medium composition	45
Bioreactors.....	45
Microtiter plates	45
Sampling	46
Enzymatic assays	46
Biolog experiments.....	47
Works Cited	49
Chapter 4. Screening and upscale	52
Rational Selection of Nitrogen Source for Optimized Heterologous Protein Production using <i>Streptomyces lividans</i>	53
Abstract	53
Introduction.....	53
Materials and Methods	55
Strains and media	55
Biolog experiments.....	55
Microtiter experiments	56
Fermentation experiments.....	56
Data handling.....	57
Results	57
Initial Screening of 95 nitrogen sources	57
Verification of physiological performance on selected nitrogen sources in 3-mL scale	58
Detailed characterisation of growth and production on selected nitrogen sources	60
Discussion	61
Nitrogen source groups	61
Scale up of experiments	64

Final remarks	65
References	65
Chapter 5. Chemostats	78
Fermentation	78
Chemostat	78
Previous studies with filamentous bacteria	80
Physiological studies	81
Strains and media	82
Fermentation experiments	82
Results	84
Discussion	89
Works Cited	91
Chapter 6. Beyond the black box	93
The black box	93
Beyond the Black box	93
Analytical approaches of looking inside the cells	95
Genetics	95
Genomics	96
Transcriptome	100
GSM	101
Proteome	104
Metabolome	105
Flux	107
Works Cited	111
Chapter 7. Discussion and perspectives	118
Summing up	118
Perspectives	119
Genomics	119
Regulation	119
Modelling and 'omics'	120
Classical vs. modern methods	120
Genetic stability	121
Fermentation conditions	122

Model strains.....	122
Concluding remarks.....	122
Chapter 8. Appendix A – Additional carbon source figures	124
Chapter 9. Appendix B – Table of contents of Biolog plates.....	127
Chapter 10. Appendix C - The draft of the paper (Rattleff, et al., 2013b).....	135
Metabolic flux profiling of sixteen <i>Actinobacteria</i> species	136
Abstract	137
Introduction.....	138
Materials and Methods	140

Chapter 1. Outline

Actinomycetes are widely recognised as producers of antibiotics and secondary metabolites with a wide range of biological activities; further they have been investigated as producers of heterologous proteins. Though, the amount of published literature within this field is very limited compared to publications regarding secondary metabolites. For a host to be considered for heterologous production of protein, i.e. to be used as protein cell factory, some basic requirements have to be in place.

- knowledge of the genomic sequence
- access to genetic tools
- knowledge about metabolism
- knowledge about physiology across scales of production
- know-how about scale-up
- knowledge about regulation
- the capacity to perform reproducibly in cultivations, also related to kinetic parameters such as
 - growth rate
 - productivity
 - substrate uptake
 - lag-phase, and experiment duration

Actinomycetes are considered promising hosts for heterologous proteins production because as a benefit of the thorough interest in secondary metabolites some of the cell factory pre-requisites have been met. A number of *Actinomycetes* have been full genome sequenced, and several more are in progress. The genetics have been thoroughly studied in a number of hosts, therefore metabolism has been described thoroughly, and general physiology is known, besides the developments of genetic tools. *Actinomycetes* also possess an inherent capacity to secrete large amounts of proteins (enzymes), as part of their saprophytic lifestyle, which has been one of the primary criteria for selecting them as potential good hosts. The importance of high secretion capacity is that exporting the product to the growth medium eases downstream processing and thereby reduces process costs.

Streptomyces lividans, specifically, is one of the particularly promising *Actinomycetes*, since this organism already has been used in industrial scale for production of human proteins (Brawner, 1994). Further this soil bacterium is recognised to have a very low endogenous extracellular protease activity (Lussier, et al., 2010) ensuring product stability, as well as improving the yield.

S. lividans has been studied as a heterologous host for more than 20 years (Anné & Mellaert, 1993) (Brawner, 1994), and during that time a number of important developments have been made, including an efficient expression system, characterisation of a number of promoters and secretion-signal peptides, molecular characterisation of the export systems (SEC and TAT), creation of a wide range of cloning vectors, and establishment of protocols for genetic and molecular work. The amount of work available on quantitative physiology in bioreactors, relating to primary metabolism however is sparse, though as a host for heterologous secondary metabolites *S. lividans* has proved its worth.

The project that is the basis for this thesis has therefore sought to address exploiting the potential of *S. lividans* as a cell factory. The work described in this thesis has been closely tied to the EU framework 6 project named STREPTOMICS that was focussed on developing *S. lividans* as a cell factory by studying the effect of heterologous protein production on the physiology of the organism, and possibly suggesting and implementing metabolic engineering optimisations to improve production. The members of the consortium included specialists in many fields, some familiar with, others experts in, *Streptomyces*, some were brought in for their expertise in a particular field, despite their lack of specific experience with *Streptomyces*.

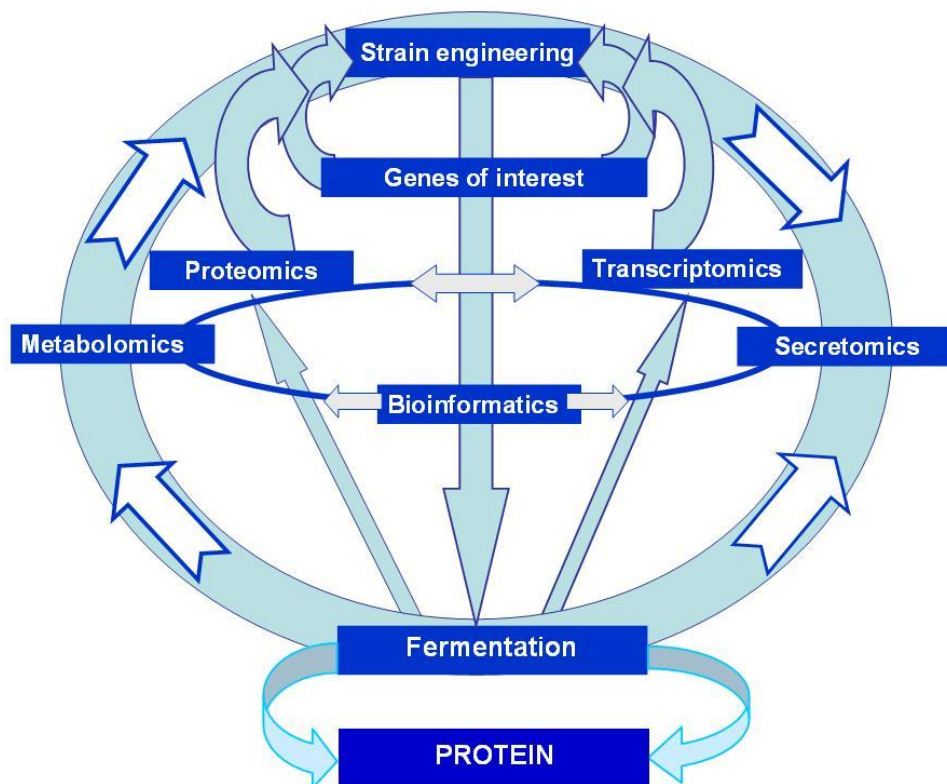


Figure 1 adapted from the technical annex of the STREPTOMICS project. This figure gives an overview of the STREPTOMICS project, and also relates how interconnected all of the operations are. Identifying genes of interest and expressing these are only the first steps. Subsequently a great deal of fermentation experiments are required to gather the samples for the advanced 'omics' analyses, which will feedback data towards strain optimisation/engineering. Because *Streptomyces* are fairly uncharacterised, a number of bioinformatics tools need to be developed as well as help investigate the genome sequence, which should be sequenced as part of this project.

STREPTOMICS took its starting point in a number of *S. lividans* strains that have already successfully expressed a number of proteins. These were to be characterised thoroughly, and samples should be taken for advanced 'omics' analyses, which the separate members in the consortium were proficient in analysing, and included among others proteomics which was to be performed by University of Greifswald, metabolomics that was to be performed by University of Manchester, and finally transcriptomics, as well as development of bioinformatics tools that were to be performed by BioXpr, Belgium. Two partners, KU Leuven (Belgium) and FORTH (Greece), were involved in characterising the secretion machinery. DTU was responsible for quantitative physiology, and flux analysis of strains, as well as obtaining samples from reproducible conditions in bioreactors, in order to more thoroughly be able to compare the results from the analyses. Genome sequencing as well as Bio-prospecting among Icelandic thermophiles for potential

products was to be performed by Matis Ltd. an Icelandic Food and Biotech R&D institute. Further, a number of small and medium enterprises (SME's) have been included in the project, with various roles, reflecting that the project has an industrial relevant angle.

As depicted in Figure 2 the work in this thesis has been shaped to facilitate the overlying EU project, and it has been a great opportunity to collaborate with prime researchers in their respective fields. The Centre for Microbial Biotechnology has had a long experience with antibiotic and secondary metabolite production in the strains *S. coelicolor* and *Amycolatopsis balhimycina*, which meant that some infrastructure was in place, though adaptation and validation of methods had to be performed. The other project partners have primarily been utilising classical rich media, therefore a defined minimal medium had to be introduced to ease quantitative and 'omics' analyses, and ensure comparability of results between the groups.

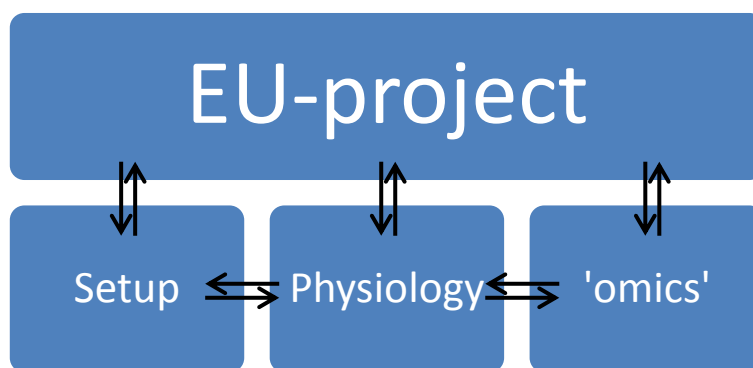


Figure 2 An overview of how the various parts of my project has evolved controlled by the needs of the EU project

The quantitative physiology was primarily considered by DTU in the EU project, since the other members were interested in developing the analyses of their core competence. Where biomass had not yet been supplied, the partners performed experiments in shake flasks, which obviously caused a higher degree of variance caused by the imperfect control of mass transfer-, as well as physio-chemical -, parameters in shake flasks.

The thesis is the result of work performed here at DTU, which has been of a developmental nature as little has been reported within this field. A lot of basic work had to be performed, both to establish working conditions, as well as to support the EU project. The limited number of published papers reflects the challenging nature of establishing a cell factory, which others hopefully can build on.

The shape the thesis takes resembles the overall shape of the EU project, as depicted in Figure 2. Before delving into the work performed, a thorough introduction into *Streptomyces lividans* as a cell factory is needed, and will be presented in chapter 2. An overview will be given of cell factories and common hosts for protein production. The setup includes generating propagation procedures, as well as screening of carbon and nitrogen sources for effect on growth and product formation. These elements are included in the chapters 3 and 4, of which the latter is a paper submitted to Applied Microbiology and Biotechnology (AMB) (Rattleff, et al., 2013a). Physiology is also the core of the chapters 3, 4 and 5, where quantitative physiology was studied across laboratory relevant scales, from microtiter plates to 1L lab scale bioreactors. Further, in chapter 5 the effect of growth rate on product formation was studied in chemostats. The chemostats with their highly controlled conditions are useful for reproducibly sampling for the advanced

'omics' analyses, and thus support the EU project. Lastly, in chapter 6 an overview will be given of what 'omics' analyses are available, what results have been obtained, and what perspectives are in the development and use of these analyses. This chapter is being supported by two papers that are appended as drafts, one is the preliminary flux analysis, to assess ED-pathway activity in a number of *Actinomycetes* (Rattleff, et al., 2013b), the other is a collaborative effort to develop an unstructured kinetic model to simulate the fermentation processes of *S. lividans* (Nandy, et al., 2013). Chapter 7 contains the concluding remarks, on the evaluation, and perspectives, of *S. lividans* as a cell factory for heterologous protein expression.

Works Cited

Anné, J. & Mellaert, L. V., 1993. *Streptomyces lividans* as host for heterologous protein production.. *FEMS Microbiol Lett*, Dec, 114(2), pp. 121-128.

Brawner, M. E., 1994. Advances in heterologous gene expression by *Streptomyces*.. *Curr Opin Biotechnol*, Oct, 5(5), pp. 475-481.

Lussier, F.-X., Denis, F. & Shareck, F., 2010. Adaptation of the highly productive T7 expression system to *Streptomyces lividans*.. *Appl Environ Microbiol*, Feb, 76(3), pp. 967-970.

Nandy, S. K., Rattleff, S., Krist, G. & Lantz, A. E., 2013. Model of enhanced heterologous protein production by *Streptomyces lividans* – Effect of C/N ratio. *Manuscript in prep*, p. In preparation.

Rattleff, S., Nielsen, J., Rokem, S. & Lantz, A. E., 2013b. Metabolic flux profiling of sixteen Actinobacteria species. *Manuscript in prep*.

Rattleff, S., Thykaer, J. & Lantz, A. E., 2013a. Rational selection of nitrogen source for optimized heterologous protein production using *Streptomyces lividans*. *Applied Microbiology and Biotechnology*.

Chapter 2. Introduction

Industrial biotechnology

The power of microbes has been harnessed over the course of the centuries. Only in the more recent history has the human race become knowledgeable about the microbes, and have begun to investigate them more in-depth. Industrially, microorganisms have seen use in the brewing and baking industry almost since the dawn of time. The use of fermented food and drink go back a long way through history, not only to increase shelf life, but also as a means to increase nutritional value (Demain, 2010). There are indications that even in ancient Rome 100 BC there were 250 bakeries selling leavened bread (Demain, 2010). Brewing was originally something performed on an individual basis. In the mid-16th century industrialization took place, where monasteries and breweries took over the brewing, and sold their product to the local pubs.

Biotechnology was a term first believed to be coined in 1919 by a Hungarian agricultural engineer Karl Ereky. Originally the term meant “all the lines of work by which products are produced from raw materials with the aid of living organisms” (Bud, 1989). Over the years the understanding of the term has changed, with the change in technology available, suffusing towards microbial biotechnology.

Industrial microbiology became interesting with the need for glycerol for production of ammunition during World War one, and later with the advent of antibiotics large scale fermentation processes were developed rapidly (Vinci, 2003). Another pioneering microbial process was the Acetone-Butanol-(later also ethanol) fermentation by *Clostridium* which is considered the first non-food fermentation (Demain, 2010). Fermentation for production of food and industrial products has the pronounced advantage that variability in production is less, and the scaling of production to fit demand is easier than using crops or purification from these. For some processes the environmental impact is also reduced compared to e.g. petro-chemical applications. Reduced variability of product was especially seen for the production of citric acid that originally was extracted from lemons, though Currie patented a process in 1916, that was later developed to industrial scale in 1923, where he used *Aspergillus niger* (Demain, 2010). Potentially the environmental impact of using bioprocesses is reduced compared to classical production methods. The waste streams will contain fewer and less toxic components.

Later, in an attempt to improve penicillin production, it was discovered that high concentrations of glucose had a detrimental effect on the production. An alternative was to use lactose that supported high productivity, but was inefficient as a growth medium. A mixed medium of glucose and lactose was a possibility, though the problem is that high glucose concentrations inhibit production of antibiotics, among other, and extending the batch by feeding small amounts of glucose, overcomes the problem of inhibition, while using a carbon source that supports growth very well, thus resulting in the invention of a fed-batch process in 1953, which is widely applied in industry today (Demain, 2010). One of the main benefits of the fed-batch is that incredibly high cell density can be obtained, and after the initial batch phase, the growth rate can be controlled by controlling the rate of addition of feed. Controlling the addition of glucose, allows e.g. to maintain an uninhibiting low concentration of glucose in the medium. This allows for a high productivity of products whether they are growth-related or -independent.

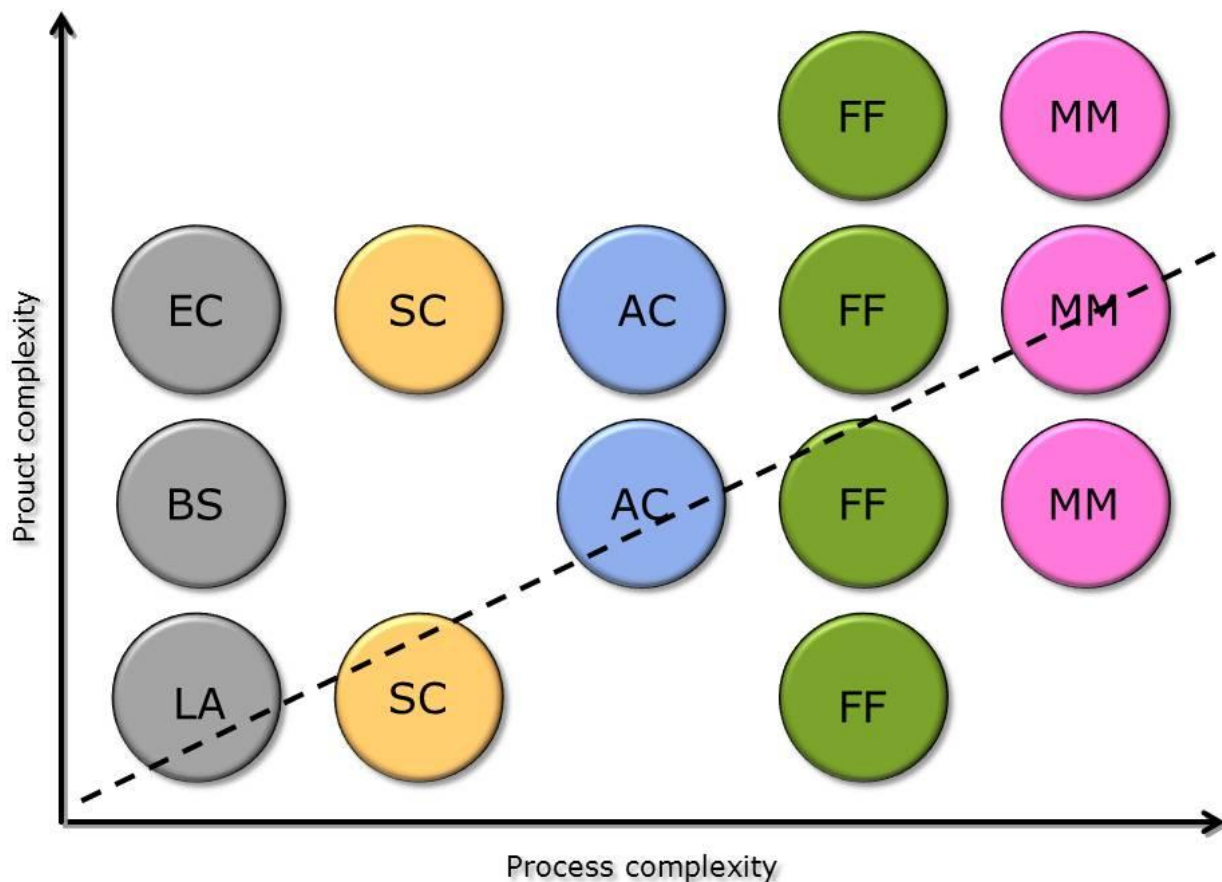


Figure 3 A schematic overview of the connection between available products, their complexity, and the complexity of the process producing them. Abbreviations are representative organisms in each category: LA (lactic acid bacteria), BS (*Bacillus subtilis*), EC (*Escherichia coli*), SC (*Saccharomyces cerevisiae*), AC (Actinomycetes), FF (filamentous fungi), MM (mammalian cells).

Figure 1 gives a schematic representation of the connection between the complexity of the process and the complexity of the product. The dashed line represents a trade-off consideration, which can be difficult to determine in reality. It is meant to suggest when it would be sensible choosing a different organism. The trade-off will be between with increasing complexity of the production and increasing complexity of the product. Some of the higher organisms allow producing more complex products, that are also higher value added, and here there is no problem with that decision. Alternatively, filamentous fungi are capable of producing high amounts of primary metabolites, though they are involved in relatively complex processes, thus the amount, and price, of the product have to be weighed accurately against the cost of running the process.

Product complexity

The complexity of the products is based on where they are derived in the metabolism, and what the requirements for using said products are. Simple products are primary metabolites, and with increasing complexity homologous proteins, secondary metabolites, simple heterologous proteins, modified secondary metabolites and lastly complex heterologous including those requiring glycosylation. More in line with choosing the organism to produce the product is the value of product in the market. Here products can be divided into bulk and high value added products. A classical high value added product

would belong in the pharmaceutical industry, with a market size estimated between 87 (2008) and 169 (2014) billion US\$, whereas the market size for heterologous (bulk industrial) proteins is estimated to be between 5.1 (2009) and 6.5 (2013) billion US\$ (Mattanovich, et al., 2012) (Erickson, et al., 2012); despite the difference in relative market values, the physical volume is not addressed. The bulk enzymes are used in almost all areas of industry e.g. feed, food, detergents, and biofuels. The distinction between bulk and high value added products is of importance when designing the process, as it implies how much of the product obtained per substrate is required for the process to be successful. Optimising bulk products are mostly related to optimising Y_{sp} , yield of product on substrate, since the cost of the substrate has a large impact on the final price of the product. In high value added products the optimization parameter is r_p , specific rate of product formation (usually g product/ (g DW * h), since the price of the substrate is less important; it is more interesting how long the process will be, what by-products there are etc. The media for high value added products can become quite expensive because of the wish for a defined product spectrum, the requirements of the organism, and bureaucratic regulatory concerns depending on whether the product is pharmaceutical.

Known primary metabolites are among others: vitamins, amino acids (food additives), organic acids, and solvents. Examples of the latter two could be citric acid, acetone, butanol, and ethanol. All of these are established bulk products (Makino, et al., 2011) (Adrio & Demain, 2006) (Soetaert & Vandamme, 2006).

Secondary metabolites have biological activities, and are often of an ecologically adaptive nature (Bibb, 2005). The activities include: antibiotics, antitumor, antimalarial, anthelmintic, and colours. One of the most famous secondary metabolites is penicillin, discovered by Alexander Fleming, and many antibiotics have since been discovered. Secondary metabolites are often found from hosts that inhabit selective niches, and have complex living requirements; while it is possible to develop each specific host for producing the compound of interest, the complex regulatory networks controlling secondary metabolites make them excellent subjects for heterologous expression. Estimated 60% of all antibiotics are produced by members of the order *Actinomycetales*, and heterologous expression of antibiotics in this order has long been an area of interest.

Homologous proteins are, in terms of complexity, considered on-par with secondary metabolites in this representation, while actually belonging to the primary metabolism. These proteins are often sought because of specific activity, and are to be produced in bulk. Production of homologous proteins depends on amount and activity and very much on the capability of the natural producer, but a wide range of natural hosts are available, including bacteria and filamentous fungi.

The more advanced products are mainly concerned with heterologous proteins which will be elucidated later, though this group can be subdivided depending on whether glycosylation is required, which shifts the process requirements towards higher complexity.

Process/organism complexity

Organism complexity and process complexity are closely related, since the increasing demands of the organisms reflect how complex the process is to handle. The bacteria and unicellular yeasts are relatively uncomplicated to use in bioprocesses, while mass and energy transfer issues are relevant to keep in mind for high density aerobic processes, their very simple morphology makes no restraints. Their product spectrum varies, but is concerned primarily with the simpler products, such as primary metabolites,

solvents (ethanol, butanol...), vitamins, amino acids, protein (single cell protein, homologous proteins, and simple heterologous proteins). Thus a vast range of processes are represented by this group of organisms. (Morello, et al., 2008) (Mattanovich, et al., 2012) (van Ooyen, et al., 2006) (Demain & Vaishnav, 2009).

In the area of higher process complexity the product spectrum expands to include the very complex heterologous proteins with human glycosylation patterns. Filamentous fungi (FF) are represented in all avenues of production, from primary metabolites (organic acids), over homologous proteins, to heterologous proteins of even moderately complex character including proteins requiring correct glycosylation to function. The mammalian cells are because of their sensitivity, complex feeding requirements, the price of their media, and their growth rate primarily used for producing high value complex proteins which require correct glycosylation. They are difficult to handle, but due to their nature capable of producing human, or mammal, proteins quite efficiently.

What lacks to be mentioned from figure 1 is that in the sweet spot, less complex than filamentous fungi, though with the capability for producing somewhat complicated products, are the actinomycetes. They have been investigated for many years as potential hosts for heterologous protein production due to their excellent innate ability to excrete proteins to their surroundings (Brawner, 1994) (Binnie, et al., 1997) (Anné, et al., 2012).

Microbial cell factories

The term “microbial cell factories” is relatively new, sprung from industrial biotechnology. It is the aim to convert a cell into a cell factory i.e. the cell should be able to produce the wanted compound, from a cheaper source of raw material. The biotransformation should preferably be competitive with alternative production methods, i.e. production should be carried out with high efficiency, a limited number and amount of by-products, and high selectivity (should the product be e.g. amino acids, or in another way stereoselective). Further, for an organism to come into consideration as microbial cell factory, genetic tools for modification should be available, the strain should be genetically stable, it is of great benefit if the genomic sequence is available, and the strain should be amenable to high cell density cultivations.

These features are very important to establish a cell factory, and it is obvious that genetic stability and process reproducibility are key factors in an industrial setting. Further, choosing a production strain that produces no toxic compounds, or even carries a US-FDA GRAS (US food and drug administration, generally regarded as safe) status for food and/or feed processes, can be highly valuable in terms of easing governmental regulatory constraints. After having introduced the changes needed for obtaining the product, and preferably in high amount, it is important also that the process can run stably for an extended duration, without losing productivity. Reproducibility and ease of handling should allow the process to run reliably and preferably in a partly automated fashion. A further feature that is wanted is product stability. A high productivity is not worth much if the product is constantly being degraded, which is often the case for protein products, since many heterologous hosts produce proteases as well.

Heterologous protein production

In nature proteins (enzymes) can be found with very unique properties. In order to use these proteins they have to be produced in sufficient amounts that their price does not exceed their applicability. Many of the proteins with interesting activities are found in hosts that are unviable for industrial production, e.g. arctic phosphatase, thermo stable proteases, lipases, and cellulases. All of these proteins have been isolated from

extremophile hosts that either requires high growth temperatures (increasing the energy consumption, thus the price for the process, or have very low growth rates, making the process unbearably long. Thus the production of heterologous proteins is achieved by specific application of microbes with desirable properties to produce a, to the host, non-native protein. The first industrially produced heterologous protein was insulin produced in *E. coli* by Eli Lilly. So far, very few pharmaceutical heterologous protein products have actually been approved by the authorities, the number lies in the lower end of hundreds (Anné, et al., 2012) (Mattanovich, et al., 2012).

Heterologous proteins can be divided into:

Bulk: industrial proteins, e.g. proteases, xylanases, cellulases, phosphatases, lipases, amylases

Specialised: e.g. restriction enzymes, polymerases

Pharmaceutical: IFN-alpha2, TNF-alpha, antibodies, human proteins (EPO)

Especially the bulk enzymes are often produced from endogenous hosts, since they already have a high capacity for expressing a given protein. There is also the possibility that this production can be further optimised, though there is a limitation, before regulatory processes become a problem. Heterologous expression is most often chosen for the more specialized products, where purification from live tissue would be impractical or connected to risk, e.g. insulin, or in cases where the natural titers are so low that it is vastly uneconomical to choose endogenous production. Further, heterologous production has the distinct advantage that a production strain can be chosen, for which there is well established fermentation know-how.

One major problem with heterologous protein production is that many of the known, and currently used, producers also have a high capacity to produce proteases. This reduces product stability, and yield. Especially the filamentous fungi, though also the gram positive *Bacillus spp.*, are known for their proteases (van den Hombergh, et al., 1997) (Westers, et al., 2004) (Punt, et al., 2002). Different ways to reduce protease levels have been applied both knock outs of protease gene, morphological “engineering” (Wang, et al., 2005). Though often the residual activity is still high enough to cause substantial product loss, or the knock-out strain may be severely hampered in its growth.

Endogenous vs. heterologous host

Choosing if a protein should be produced in an endogenous or heterologous host relies much upon the characteristics outlined in the previous sections. Is it a bulk product, and is the endogenous host capable of producing sufficient amounts of the product, under conditions and with media components that will not cause regulatory concerns. If it is a homologous product, and it is already produced as one of the major products of the cell, then only culturing demands can bar the way for using the homologous strain. If however, the product is a minor product, or originally produced from a host, e.g. a plant that cannot easily be cultured to yield sufficient product, the choice of heterologous production is clear.

Well established heterologous hosts for protein production

Bacterial hosts

Escherichia coli is probably the most well-known host. It is very well studied, and a large repertoire of genetic tools is available. *E. coli* grows fast; under favourable conditions a generation time can be as low as 20-30 min. It is easy to handle, and supports high cell density fermentation, up to 150 g/L dry weight (Schmidt, 2004). This in conjunction with the absence of secondary metabolites makes *E. coli* a good host. It is a gram negative, which means it has a cell wall, besides the cellular membrane. This has a very profound effect on this organism as a host for heterologous protein production. Many proteins are accumulated, either in the cytoplasm, or in the periplasmic space, as inclusion bodies. This is problematic, since the aggregates are inactive protein that needs to be refolded to obtain functional activity. Though proteins that require sulphide bonds can be accumulated in the periplasmic space, and subsequently released, after subjecting the culture to an osmotic shock (Terpe, 2006). Some researchers have also increased yield by co-expressing chaperones, to help fold the product correctly. There is mention in the literature of intracellular and periplasmic proteases, and subsequently protease deficient mutants, though *E. coli* is generally recognized as having low protease activity. Historically the first available commercial heterologous product was insulin produced in *E. coli* by Eli Lilly, and many products have followed. As much as 30% of the FDA approved products are produced in *E. coli* (Huang, et al., 2012).

Bacillus spp. and *Lactococcus lactis* are gram positives and have high capacities for excreting active protein into the growth medium. This is favourable in terms of downstream processing, making purification easier. *Bacillus subtilis* is almost as well studied, and well known as *E. coli*. *Bacillus* is industrially a well-tested genus, and around 50-60% of industrial enzymes are produced in *Bacillus spp* (Schallmey, et al., 2004) (Westers, et al., 2004). The genomic sequences of a number of bacilli are available, as are a large number of genetic tools. Terpe claims it is more difficult to transform bacilli, since protoplastation is necessary (Terpe, 2006), though compared to working with filamentous fungi this is likely a small encumbrance. Unfortunately many gram positives have an inherent capacity for producing proteases, which results in reduced product stability. Studies have been conducted to knock out the various proteases, with varying degree of success (Westers, et al., 2004), as well as using the less common relatives with an inherent lower protease activity. Yields (titers) in the range 3 g/L have been reported using *B. subtilis* as host, and reports have also mentioned that up to 30% of the soluble protein was heterologous (Terpe, 2006).

Lactococcus lactis is one of the new contenders for heterologous expression. Industrially it is used in the dairy industry. It has a simple product spectrum mostly concerned with primary metabolites, such as lactic acid and there is only one extracellular protein at detectable levels. Recently some work has gone into presenting *L. lactis* as hosts for heterologous protein expression, and since it only has one extracellular protease that can be inactivated, seemingly without hampering growth (Morello, et al., 2008). Further, it is a well-studied organism with a simple metabolism that supports aerobic and anaerobic growth. *L. lactis* is considered GRAS, and therefore suitable for producing enzymes to be used in food and feed industries. *Lactococcus* is well known fermentation wise and large scale fermentations are in place.

Yeast hosts

The simplest eukaryotic cells are the yeasts, which while in essence fungi, have their own group in terms of production hosts, and their mostly unicellular lifestyle. They are also one of the most studied groups of

organisms, together with the bacteria; partly due to the ease of handling, and the fact that they are eukaryotic, so they can in some instances be used as simple models of higher organisms. The most common yeast hosts are *Saccharomyces cerevisiae* and *Picchia pastoris*. Of these, the latter is known for very high protein expression levels using the highly inductive methanol promoter system. The yeasts possess the post translational machinery of eukaryotes for e.g. glycosylation that bacteria lack; though the post translational activity is only somewhat successful. *S. cerevisiae* has been shown to hyper glycosylate heterologous proteins, and uses a non-mammalian mannose glycosylation pattern (Demain & Vaishnav, 2009). Work has gone into producing humanised-yeasts, with less aggressive glycosylation patterns (van Ooyen, et al., 2006). Both of the model yeasts can be used in very high density fermentations, surpassing 200 g DW/L (Demain & Vaishnav, 2009). And they are very well studied, characterised and easy to handle both in a fermentation sense, but also genetically, e.g. *Kluyveromyces lactis* was the first yeast for which a transformation system was introduced (van Ooyen, et al., 2006). Many of the yeasts are considered GRAS, since they have been used throughout history (Juturu & Wu, 2012), among these *K. lactis*, which industrially has been used to produce both lactase and later bovine chymosin, the first higher eukaryotic protein expressed in a microorganism (van Ooyen, et al., 2006). Yeasts have a proven track record in terms of large scale industrial production, also of heterologous proteins, with the added benefit that for the most part, they are able to secrete fully folded protein. Protease activity is a known issue, and for *Picchia* protease knock-out mutants have been produced (Macauley-Patrick, et al., 2005). Unfortunately, the methylotrophic yeasts, including *Picchia*, that use the highly inductive methanol promoter system, require explosion proof fermentation equipment, which increases the capital investment when setting up the process (van Ooyen, et al., 2006). While yeasts are easy to handle in terms of physiology, they produce a lot of heat and CO₂, and for aerobic processes need a high aeration, resulting in high energy consumption.

In cases where the yeast is Crabtree positive, then the process will draw out, due to the production of ethanol at high growth rates, that later has to be taken up again, and metabolised, at a much slower growth rate. To prevent Crabtree effect in continuous or fed-batch processes the feeding rate has to be below the critical substrate uptake rate, thus growth has to be kept below 0.1 h⁻¹ (Mattanovich, et al., 2012), off-setting one the yeasts main benefits, a reasonably fast growth rate.

Fungal hosts

Fungi are recognized to have a very developed capacity for excreting active product, by merit of their saprophytic lifestyle. Homologous products are known in titers of 10-30 g/L e.g. proteases and amylases (Wang, et al., 2005) (Punt, et al., 2002), though heterologous products are nowhere near this scale yet; many fungi produce high amounts of secreted proteases reducing product yield (van den Hombergh, et al., 1997). Much effort has gone into characterizing and minimizing the extracellular protease production, in heterologous protein fermentations. It has been found that pH, inoculum, medium composition and morphology are some of the features that affect protease production. One of the approaches to reduce extracellular protease production is to make genetic knock-outs. This approach, while somewhat successful (Punt, et al., 2002), also has caused the mutants to be debilitated to a degree that makes them undesirable for large scale production (Wang, et al., 2005). And still there is protease activity, even if only 1-2% of the parental strain, it is sufficient for degrading heterologous proteins (Mattern, et al., 1992). Even knock-outs of transcription factors controlling proteases show limited success (Punt, et al., 2008).

In the recent past much research focus has been put into creating genetic tools for expanding the toolbox of filamentous fungi. Despite the expanding toolbox, knowledge of the secretory pathway is lacking compared to bacterial systems (Wang, et al., 2005) (Mackenzie, et al., 1987). Due to some court cases, concerns over intellectual property matters regarding the use of fungi have increased the interest in expanding the range of available hosts; this expansion is also with the consideration for what strains are recognized as GRAS in the food industry (Punt, et al., 2002). Two of the common genera of well-known production strains are *Aspergillus* and *Trichoderma*, known for their endogenous proteins such as proteases and cellulases and organic acid production.

One thing that characterizes filamentous fungi is their hyphal growth, which can be both of merit and detriment during cultivation depending on the product. The hyphal growth results in highly viscous fermentation broths that, depending on the particular morphology, can have non-Newtonian properties. This causes problem with mass transfer, which in turn can affect product formation. Much effort has gone into describing and controlling morphology, and estimating the impact of morphology on product formation. It is a complex problem, since it can be difficult to discern whether the effect seen is caused by the physiological or morphological change. Some researchers have reported that it seems the variability in productivity is caused mainly by physiology, less by morphology, as long as the morphology does not impart mass transfer limitations (Wang, et al., 2005). The most used process when it comes to filamentous fungi is the submerged fermentation, and while great know-how is available, recent developments have given increased impetus to developing solid state fermentation, as the genetic regulation here is vastly different (Nevalainen, et al., 2005), and potentially novel products can be developed, that are unavailable or produced in low amounts in submerged fermentations (Oda, et al., 2006). Despite their great promise, so far no heterologous pharmaceutical products are produced by filamentous fungi (Wang, et al., 2005); though their presence in industrial fermentations is endemic.

Streptomycetes – state of the art

Streptomyces are gram positive soil bacteria, and have long been recognized for their plethora of secondary metabolites, which have been one of the main driving forces for developing the *Streptomyces* genetics. It is a very diverse group spanning many habitats, and many *Streptomyces* can be found possessing very unique abilities, further increasing interest in studying them, both for their potential products, but also their capabilities as heterologous hosts for these secondary metabolites. Not only are they capable of producing interesting secondary metabolites, by virtue of their saprophytic lifestyle, *Streptomyces* are specialised in degrading macro cellular substrates, and therefore both capable of producing proteins with unique abilities but also secreting high amounts of proteins. Industrially actinomycetes have been used to produce many secondary metabolites, including antibiotics, and a number of industrially relevant, and a few pharmaceutical, proteins. Specifically Leucotropin™ by Cangene is produced using *S. lividans* as host (Brawner, 1994) (Binnie, et al., 1997) (van Wezel, et al., 2006). What makes *S. lividans* a promising host for heterologous protein production, compared to some other actinomycetes, aside from high secretion efficiency is the surprisingly low extracellular protease activity (Strickler, et al., 1992) (Anné, et al., 2012). In terms of genetics, *S. lividans* is also favoured, since it has become the model for heterologous protein studies of *Streptomyces* due to a very relaxed restriction system, and the fact that it does not methylate its own DNA, easing transformation. The close relative *S. coelicolor* A3(2) is on the other hand a favoured in studies of genetics and regulation due to the coloured antibiotics it produces (Kieser, et al., 2000). The two strains are so genetically closely related that expression arrays from *S. coelicolor* A3(2) has been used for *S.*

lividans (Jayapal, et al., 2007), and in the comparative genomics investigation it was found that *S. lividans* differs from *S. coelicolor* mainly in laterally acquired genetic elements (Lewis, et al., 2010) (Jayapal, et al., 2007). *S. coelicolor* A3(2) has been full genome sequenced (Bentley, et al., 2002), and while several sequencing initiatives have been performed, an annotated full genome sequence is still lacking for *S. lividans*, though the Broad institute has the partial genomic sequence (draft) of *S. lividans* among several actinomycetes¹. Genomic sequence can therefore be considered accessible, if not completely known. A host of genetic tools are available (Kieser, et al., 2000) (Brawner, 1994) (Vrancken, et al., 2010), and detailed knowledge about the metabolism is being gathered (Hodgson, 2000). *S. lividans* only has a limited spectrum of secondary metabolites, and the before mentioned low exogenous protease activity, making it a good choice for heterologous expression (Jayapal, et al., 2007) (Kieser, et al., 2000) (Anné & Mellaert, 1993) (Vrancken & Anné, 2009).

Using *Streptomyces* as heterologous hosts have been considered many times over the last few decades (Brawner, 1994) (Hong, et al., 2003) (Vrancken & Anné, 2009), and in the cases of (Brawner, 1994) (Vrancken, et al., 2007) (Schaerlaekens, et al., 2001) (Mellaert, et al., 1998) among others, effort has been made developing the molecular and genetic tools for using *Streptomyces* as heterologous hosts. An example of a molecular genetic tool used for large scale manipulation of the genome is based on the use of PCR-targeted substrates flanked by short direct repeats. The target sequences are created from a Supercos-1 library, and the gene can be deleted, in-frame, by a double cross-over event (Gust, et al., 2003). The method represents establishing the use of technology moved between different organisms, as both *E. coli* and *S. cerevisiae* methodology is applied to improve speed and efficiency of genetic manipulations in *Streptomyces*.

Despite the present knowledge of metabolism, most of the know-how of fermentation in larger scales is kept with companies, and precious little has been published. Brawner and Binnie are affiliated with Cangene, a company that uses *S. lividans* as a production host, and they have published some papers. Otherwise, the knowledge of handling *S. lividans* has been presented by Van Wezel, and very recently D'Huys (van Wezel, et al., 2006) (D'Huys, et al., 2011) (D'Huys, et al., 2012). Anné published a review in 2012 (Anné, et al., 2012), where he outlined the state of *S. lividans*, and some of the tasks that should be performed, in order to gain sufficient understanding of the *S. lividans* biology to introduce *S. lividans* as a general heterologous production host next to the classical hosts, *E. coli* and *Bacillus*. Most of the features in this review are concerned with physiological knowledge, Flux, Metabolomics, Proteomics etc., and molecular knowledge of the protein transport systems, SEC and TAT.

Very little has been published on the use of *S. lividans* in chemostats (Rossa, et al., 2002), aside from some very thorough genetic stability studies in the 80's and 90's. Concerning heterologous proteins a few studies can be found, which are mainly studying alternative limitations, PO₄, Mg²⁺ and carbon catabolite repression, more on this in chapter 4.

Outline of this thesis

The aim of this thesis has been to investigate the physiology of heterologous protein producing strains of *S. lividans*.

¹ http://www.broadinstitute.org/annotation/genome/streptomyces_group/GenomeStats.html accessed 16-03-2013

Despite much knowledge on the subject of general metabolism, we would like to know more specifically what the impact is of various carbon and nitrogen sources are on *S. lividans* as production host for heterologous protein production.

Further, I wish to study the physiology of *S. lividans* in liquid fermentation in different laboratory related scales and establish chemostats, where the impact of the growth rate on the production of heterologous proteins is being estimated.

Lastly, based on the physiological studies, the next step is to expand the knowledge, both to develop a model of *S. lividans*, that can predict growth and product formation, and verify this model with fermentation data; as well as we wish to be able to sample biomass for advanced 'omics analyses, that can help deepen our understanding of the intracellular reaction networks and regulation of *S. lividans*.

Works Cited

- Adrio, J. L. & Demain, A. L., 2006. Genetic improvement of processes yielding microbial products.. *FEMS Microbiol Rev*, Mar, 30(2), pp. 187-214.
- Anné, J. et al., 2012. Recombinant protein production and streptomycetes.. *J Biotechnol*, Apr, 158(4), pp. 159-167.
- Anné, J. & Mellaert, L. V., 1993. Streptomyces lividans as host for heterologous protein production.. *FEMS Microbiol Lett*, Dec, 114(2), pp. 121-128.
- Bentley, S. D. et al., 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).. *Nature*, May, 417(6885), pp. 141-147.
- Bibb, M. J., 2005. Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*, 8(2), pp. 208-215.
- Binnie, C., Cossar, J. D. & Stewart, D. I., 1997. Heterologous biopharmaceutical protein expression in Streptomyces.. *Trends Biotechnol*, Aug, 15(8), pp. 315-320.
- Brawner, M. E., 1994. Advances in heterologous gene expression by Streptomyces.. *Curr Opin Biotechnol*, Oct, 5(5), pp. 475-481.
- Bud, R., 1989. Janus-faced biotechnology: an historical perspective. *Trends in Biotechnology*, 7(9), pp. 230-233.
- Demain, A. L., 2010. History of Industrial Biotechnology. In: *Industrial Biotechnology*. s.l.:Wiley-VCH Verlag GmbH & Co. KGaA, pp. 17-77.
- Demain, A. L. & Vaishnav, P., 2009. Production of recombinant proteins by microbes and higher organisms.. *Biotechnol Adv*, 27(3), pp. 297-306.
- D'Huys, P.-J. et al., 2011. Amino acid uptake profiling of wild type and recombinant Streptomyces lividans TK24 batch fermentations.. *J Biotechnol*, Apr, 152(4), pp. 132-143.
- D'Huys, P.-J. et al., 2012. Genome-scale metabolic flux analysis of Streptomyces lividans growing on a complex medium.. *J Biotechnol*, Sep, 161(1), pp. 1-13.
- Erickson, B., Nelson & Winters, P., 2012. Perspective on opportunities in industrial biotechnology in renewable chemicals.. *Biotechnol J*, Feb, 7(2), pp. 176-185.
- Gust, B. et al., 2003. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin.. *Proc Natl Acad Sci U S A*, Feb, 100(4), pp. 1541-1546.
- Hodgson, D. A., 2000. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria.. *Adv Microb Physiol*, Volume 42, pp. 47-238.
- Hong, B., Li, Y., Mellaert, G. V. & Anné, J., 2003. Expression of biologically active human tumor necrosis factor beta by Streptomyces lividans.. *Yi Chuan Xue Bao*, Mar, 30(3), pp. 209-214.

- Huang, C.-J., Lin, H. & Yang, X., 2012. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements.. *J Ind Microbiol Biotechnol*, Mar, 39(3), pp. 383-399.
- Jayapal, K. P. et al., 2007. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*.. *BMC Genomics*, Volume 8, p. 229.
- Juturu, V. & Wu, J. C., 2012. Microbial xylanases: engineering, production and industrial applications.. *Biotechnol Adv*, Nov, 30(6), pp. 1219-1227.
- Kieser, T. et al., 2000. *Practical streptomyces genetics*. s.l.:John Innes Foundation.
- Lewis, R. et al., 2010. Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. *BMC Genomics*, 11(1), p. 682.
- Macauley-Patrick, S., Fazenda, M. L., McNeil, B. & Harvey, L. M., 2005. Heterologous protein production using the *Pichia pastoris* expression system.. *Yeast*, Mar, 22(4), pp. 249-270.
- Mackenzie, C. R., Bilous, D., Schneider, H. & Johnson, K. G., 1987. Induction of Cellulolytic and Xylanolytic Enzyme Systems in *Streptomyces* spp.. *Appl Environ Microbiol*, Dec, 53(12), pp. 2835-2839.
- Makino, T., Skretas, G. & Georgiou, G., 2011. Strain engineering for improved expression of recombinant proteins in bacteria.. *Microb Cell Fact*, Volume 10, p. 32.
- Mattanovich, D. et al., 2012. Recombinant protein production in yeasts.. *Methods Mol Biol*, Volume 824, pp. 329-358.
- Mattern, I. E. et al., 1992. Isolation and characterization of mutants of *Aspergillus niger* deficient in extracellular proteases.. *Mol Gen Genet*, Aug, 234(2), pp. 332-336.
- Mellaert, L. V. et al., 1998. Molecular characterization of a novel subtilisin inhibitor protein produced by *Streptomyces venezuelae* CBS762.70.. *DNA Seq*, Mar, 9(1), pp. 19-30.
- Morello, E. et al., 2008. *Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion.. *J Mol Microbiol Biotechnol*, 14(1-3), pp. 48-58.
- Nevalainen, K. H., Te'o, V. S. & Bergquist, P. L., 2005. Heterologous protein expression in filamentous fungi. *Trends in Biotechnology*, 23(9), pp. 468-474.
- Oda, K. et al., 2006. Proteomic analysis of extracellular proteins from *Aspergillus oryzae* grown under submerged and solid-state culture conditions.. *Appl Environ Microbiol*, May, 72(5), pp. 3448-3457.
- Punt, P. J. et al., 2008. Characterization of the *Aspergillus niger* prtT, a unique regulator of extracellular protease encoding genes.. *Fungal Genet Biol*, Dec, 45(12), pp. 1591-1599.
- Punt, P. J. et al., 2002. Filamentous fungi as cell factories for heterologous protein production.. *Trends Biotechnol*, May, 20(5), pp. 200-206.

- Rossa, C. A. et al., 2002. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*.. *Metab Eng*, Apr, 4(2), pp. 138-150.
- Schaerlaekens, K. et al., 2001. Twin-arginine translocation pathway in *Streptomyces lividans*.. *J Bacteriol*, Dec, 183(23), pp. 6727-6732.
- Schallmey, M., Singh, A. & Ward, O. P., 2004. Developments in the use of *Bacillus* species for industrial production.. *Can J Microbiol*, Jan, 50(1), pp. 1-17.
- Schmidt, F. R., 2004. Recombinant expression systems in the pharmaceutical industry.. *Appl Microbiol Biotechnol*, Sep, 65(4), pp. 363-372.
- Soetaert, W. & Vandamme, E., 2006. The impact of industrial biotechnology.. *Biotechnol J*, 1(7-8), pp. 756-769.
- Strickler, J. E. et al., 1992. Two novel *Streptomyces* protein protease inhibitors. Purification, activity, cloning, and expression.. *J Biol Chem*, Feb, 267(5), pp. 3236-3241.
- Terpe, K., 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.. *Appl Microbiol Biotechnol*, Sep, 72(2), pp. 211-222.
- van den Hombergh, J. P., van de Vondervoort, P. J., Fraissinet-Tachet, L. & Visser, J., 1997. *Aspergillus* as a host for heterologous protein production: the problem of proteases.. *Trends Biotechnol*, Jul, 15(7), pp. 256-263.
- van Ooyen, A. J. J. et al., 2006. Heterologous protein production in the yeast *Kluyveromyces lactis*.. *FEMS Yeast Res*, May, 6(3), pp. 381-392.
- van Wezel, G. P. et al., 2006. Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering.. *Appl Environ Microbiol*, Aug, 72(8), pp. 5283-5288.
- Vinci, V. A. a. P. S. R. ed., 2003. *Handbook of Industrial Cell Culture: Mammalian, Microbial, and Plant Cells*. s.l.:Humana Press.
- Vrancken, K. & Anné, J., 2009. Secretory production of recombinant proteins by *Streptomyces*.. *Future Microbiol*, Mar, 4(2), pp. 181-188.
- Vrancken, K. et al., 2007. *pspA* overexpression in *Streptomyces lividans* improves both Sec- and Tat-dependent protein secretion.. *Appl Microbiol Biotechnol*, Jan, 73(5), pp. 1150-1157.
- Vrancken, K., Mellaert, L. V. & Anné, J., 2010. Cloning and expression vectors for a Gram-positive host, *Streptomyces lividans*.. *Methods Mol Biol*, Volume 668, pp. 97-107.
- Wang, L., Ridgway, D., Gu, T. & Moo-Young, M., 2005. Bioprocessing strategies to improve heterologous protein production in filamentous fungal fermentations.. *Biotechnol Adv*, Mar, 23(2), pp. 115-129.

Westers, L., Westers, H. & Quax, W. J., 2004. Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism.. *Biochim Biophys Acta*, Nov, 1694(1-3), pp. 299-310.

Chapter 3. Screening and growth of streptomycetes

Setting up a strain as a cell factory requires a great deal of knowledge of the strain and ability to control it, in terms of both reproducibility as well as genetic manipulation. One of the most basal requirements for successfully using a strain as a cell factory is that it should be predictable in the production environment; which also requires reproducible kinetic parameters.

Requirements for a cellular factory includes:

- knowledge of the genomic sequence
- access to genetic tools
- knowledge about metabolism
- knowledge about physiology across scales of production
- know-how about scale-up
- knowledge about regulation
- the capacity to perform reproducibly, also related to kinetic parameters
 - growth rate
 - productivity
 - substrate uptake
 - lag-phase, and experiment duration

Historically most studies have been concerned with cultivating *Streptomyces* either for producing antibiotics, or in smaller scales, e.g. shake flasks or on plates for genetic studies. There are a number of studies that relate to using *S. lividans* for protein production specifically in larger scales, 10L and up (Fornwald, et al., 1993) (Sianidis, et al., 2006) and (van Wezel, et al., 2006). Many more studies are available for antibiotics, and while antibiotic production may seem unrelated to heterologous protein expression there can be beneficial developments in tools and know-how that can easily be transferred. Many tools, such as genetic manipulation, cosmid libraries (Wohlleben, et al., 2009), promoter characterisation for heterologous expression (Hopwood, 1999), transformation techniques including protoplast formation (Kieser, et al., 2000), and thorough fermentation experience has become available because of this interest (Fazeli, et al., 1995) (Ives & Bushell, 1997) (Kirk, et al., 2000). When one wants to apply knowledge gained from these studies to protein production applications, it must be kept in mind that secondary metabolism usually is triggered by using stress-inducing limitations of e.g. phosphate (Bruheim, et al., 2002), and the effect of the limitation must be considered. However, important insight into the primary metabolism may still be gained from publications on antibiotic production.

Much molecular work has been done, and a number of genetic tools are available (Kieser, et al., 2000) (Vrancken, et al., 2010); more is still to come, which in term will become beneficial. General metabolic knowledge is available, and constantly being improved upon (Kieser, et al., 2000) (Hodgson, 2000). Genomic sequences are slowly being published (Omura, et al., 2001) (Bentley, et al., 2002) (Barbe, et al., 2011) (Grubbs, et al., 2011). Industrial know-how is present, but available literature on the subject is

limited. Brawner and Binnie are directly affiliated with Cangene, a company utilising *S. lividans* for production, and have been publishing some of their results and methods; otherwise the vast majority of the literature is academic. What is lacking the most is a series of physiological studies that are required to properly investigate the effects of the developed tools, and the effect on the strain of producing the products.

Streptomyces lividans, like all *Streptomyces*, are recognised by their complex developmental and extracellular biology (Flärdh & Buttner, 2009) (Chater, et al., 2010). Even more interest has been due to their wide array of secondary metabolites, which is particularly pronounced among the members of the *Streptomyces* genus (Ventura, et al., 2007). The core group of *Actinomycete* specialists at the John Innes Center have over the years assembled several guides on how to handle actinomycetes, and how to study their genetics. The first manual was published as a “Genetic Manipulation of *Streptomyces*: a laboratory manual” in 1985, followed by the first edition of the “Practical *Streptomyces* genetics”, published in the year 2000 (Kieser, et al., 2000). This guide serves as a very good starting point, and a lot can be learned about the specific *Streptomyces* biology, and a collection of general laboratory manuals are included. A good foundation in the biology is required when establishing a cell factory, which entails all levels of handling, from establishing a cell bank, developing inoculation- and sporulation procedures, establishing media compositions and sample procedures. Many of these, often time-consuming, parameters are often overlooked when comparing work between known and less well characterised strains.

Streptomyces are characterized by their very intricate growth cycle. Initially the spore germinates, and a small germ tube is formed. This is the initial mycelium, which extends to form vegetative hyphae and begins to branch. The vegetative hyphae have few septations, and possess multiple copies of the genome. In liquid culture no aerial mycelium is formed, and very few strains have the capacity to sporulate submerged. On solid culture, the vegetative mycelium, also called the substrate mycelium, extends through the substrate, and when nutrients have been mostly expended, aerial hyphae are formed (multi nucleated, reproductive mycelium (Manteca, et al., 2010)). The aerial hyphae extend upwards, begins to curl, which occurs at the same time as separation of the chromosomes occur, followed by septation, formation of the thick spore cell wall, and finally the rodlet layer that mediates hydrophobicity to the spores (Claessen, et al., 2002), which remain attached in chains. While submerged culture does not result in sporulation, the secondary metabolism can still be triggered, and ample antibiotics produced (Kirk, et al., 2000) (Borodina, et al., 2005). The filamentous mode of growth, may in liquid culture result in pellet or clump formation (Whitaker, 1991), which can create morphological conditioned mass transfer problems. Depending on the product, or purpose of the experiment, the complex morphology can introduce heterogeneity in the culture, and complicate sampling and data analysis. Morphology is an area of study for the *Streptomyces* (Whitaker, 1991), as well as for filamentous fungi, and some knowledge transfer can take place despite that the filamentous fungi are eukaryotic.

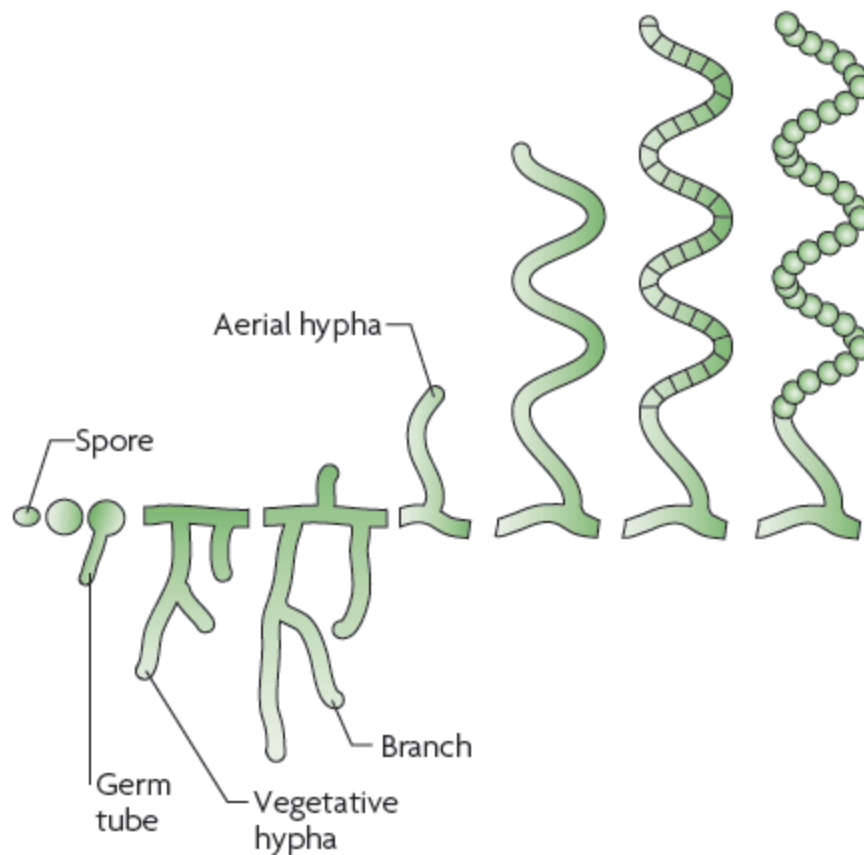


Figure 4 - a reproduction of figure 1 from (Flärdh and Buttner 2009). The developmental stages in the growth cycle of *S. coelicolor*. In liquid media most *Streptomyces* do not sporulate, and therefore do not form aerial hyphae. The formation of aerial hyphae is the beginning of a complex regulatory shift of metabolism, and the visible change was prompting much of the early genetic research, characterizing the bald genes (*bld*, no aerial hyphae), white genes (*whi*, no grey spore pigment formation). It is also during this phase that the various coloured antibiotics are produced.

To handle *Streptomyces* in liquid culture and obtain dispersed biomass, the classical methods are, for shake flask cultivations, to add physical objects to the medium that can create turbulence to increase oxygen transfer, and improve morphology. Of physical objects to have been used, steel springs (Kieser, et al., 2000), and glass beads, 20-30 beads per shake flask (Sohoni, et al., 2012) can be mentioned. In the larger scales, where physical objects would damage the bioreactor, the medium can be supplemented by additives, like surfactants, which have proven to improve morphology. Other measures to control morphology in the larger scales are possible, and inoculation concentration does have a significant influence on the process (Wang, et al., 2005), as does the physical parameters used in the bioreactor, including pH and stirrer speed (Celler, et al., 2012) (van Wezel, et al., 2006) (Wang, et al., 2005). Some additives however can be consumed by the host or complicate sampling by disturbing OD or DW, which is likely undesirable in minimal media in the academic setting, where the characterisation of the strain often is what is sought. Despite this, antifoam can hardly be avoided, since the impact of excessive foam, including the risk of contamination, loss of medium and biomass, and mixing related mass transfer problems, will likely have a deleterious effect on the experiment. Morphological engineering is not entirely trivial, as many parameters influence the process, including inoculation procedure and inoculum concentration; therefore to establish an efficient and reproducible method of working with *Streptomyces* with regards to heterologous protein production, the basic methods have to be established; the work I

have done in this regard will be the focus of the following sections. Further on a comparison between the strains grown in batch, and lastly, an example of a screening method for testing the effect of different carbon sources on protein production will be presented.

Establishing a propagation procedure

Having generated a spore stock, the next step is determining a robust and reproducible way of generating an inoculum. In the following the establishing of an inoculation procedure is described, and while generating a spore stock will not be mentioned in detail, it has been part of the preparatory experiments, and is time consuming. For storage of the strains, as well as for the stock from which the cultures will be started, two options are available. One is using spores, thus letting the strain expend the nutrients available, and complete the life cycle, producing spores, which are generally quite robust, and inert; spores can survive several freeze-thaw cycles, as long as they have not started to germinate. Creating a spore stock is time consuming, and because the spores have to germinate before they will grow, it also increases the time of the separate experiments. The other option is storing cells harvested in mid-exponential growth; this option is necessary if working with strains without sporulation capacity, but can be used by most strains. For filamentous organisms, an extra step is included in the protocol of crushing the cells in a mortar, to obtain the most dispersed stock. A frozen mycelium stock is more sensitive to storage, and failure of the same, as well as with the increased amount of handling steps, there is an increased risk of contamination. The frozen stock can be directly used as inoculum, which shortens the overall process time, though with the risk of variability in viable cell count. All of the strains I have worked with retain sporulation capacity, and therefore I have used the spore stock.

One of the first things necessary when using spores is ensuring that they germinate, preferably as synchronised as possible. Secondly, using a relatively uniform amount of viable cells, or germinated spores, will improve reproducibility of the experiments. Studies have shown that the initial concentration of cells or germinated spores in the bioreactor very much affects the physiology in the batch experiments; (Wang, et al., 2005) and (Papagianni, 2004) report that the inoculum concentration affects morphology, and morphology affects the amount of protease activity in the end of the experiment for *Aspergillus* spp. Therefore establishing a germination and inoculation procedure is necessary. Lynne Dobson suggests that a spore concentration of $1 \times 10^{7-9}$ spores/L gives smaller pellets; from 1×10^8 spores/L the major amount of the pellets is less than 0.5mm in size (Dobson, 2008), figure 3.7). (Whitaker, 1991) relays that for *S. coelicolor* antibiotic production only takes place at initial spore concentrations higher than 1×10^8 spores/L, and that in a range 3.7×10^7 – 8.7×10^8 spores/L the biomass yield is proportional to the spore concentration. I made a qualitative test on the length of the lag-phase using spore concentrations $1 \times 10^{6,7,8, \text{ and } 9}$, and found that 1×10^8 spores/L gave a short lag phase, and good growth. Pellets were visible at all concentrations tested, and the amount of the working stock required to obtain that spore concentration was balanced in terms of the concentration of the spore stock. The inoculum of 1×10^8 spores/L results in an initial OD₆₀₀ measurement of approx. 0.01-0.03. Subsequently it was necessary to determine if a special germination medium is required, or if it is possible to inoculate with spores directly in the minimal medium.

Germination

In-house there is experience with filamentous fungi, which can be germinated in the minimal medium in the bioreactor, starting with a low pH and keeping stirrer speed and gas flow low, to prevent the hydrophobic spores from being bubbled out of the medium and causing wall growth. Germination methods

of *Streptomyces* spores, according to (Kieser, et al., 2000), include heat shock or nutrient dependent germination. Nutrient dependent germination should be understood as subjecting the spores to germination inducing nutrients, often found in rich media; in the following I used the nutrient method, germinating the spores in the rich 2YT medium for a specified amount of time before transferring them to the actual growth medium.

To test whether it was possible to efficiently germinate the spores directly in a minimal medium, a comparison was performed in square deepwell microtiter plates, from EnzyScreen, with 3mL working volume, according to the method of (Sohoni, et al., 2012). The minimal medium 'germination' step (direct inoculation) was performed directly in the plates in the growth medium, setting the shaking table at a slow rotation speed (100 RPM with 20mm throw), for the same duration as the separate germination step in the rich medium. After the germination step, the directly inoculated plates were transferred to another shaker with the rotational speed turned up to growth conditions, 150 rpm with 50mm throw; the plate with the separately germinated spores was placed in the same shaker. The separate germination step took place in 50mL of the rich 2YT medium in shake flasks with 20-30 4mm glass beads. When the germination time had passed, the rich medium was transferred, without the glass beads, to 50mL Falcon tubes, and was centrifuged at 4000g for 8 min at room temperature. The supernatant was discarded, and the spores were resuspended in 5 mL sterile water, and all spores were added to the growth medium. Thereafter the growth medium was dispensed into the plate, which was then placed in the shaker at 30°C and 150 rpm with 50mm throw. The specific germination time used was 6 hours, as determined below. The spore concentration was obtained by using a specified amount of spore stock, which had previously had the spore concentration determined by CFU counting several dilutions.

In Figure 5 the two compared conditions, i.e. direct inoculation, and separate germination in rich medium, are displayed. The microscopy pictures represent samples taken from the subsequent experiment in minimal medium. Samples from each culture were taken at T=17h and T=21h. From previous experience it was possible to verify growth after around 20 hours, as this would be the time when the experiments enter into exponential growth phase, and can be sampled regularly. At T=17h in Figure 4 it can be observed that the spores that were not separately germinated, i.e. directly placed in the minimal growth medium, have not germinated. No spores can be found that display a germ tube, or vegetative hyphae, suggesting that the spores are still dormant. The spores that were separately germinated in the rich 2YT medium seem to have mostly germinated, and many vegetative hyphae are found, some have even started branching.

At T=21h for the spores that were not separately germinated, still no vegetative hyphae are found, and no spores sprouting a germ tube were found in the screened samples. For the spores separately germinated in rich medium, clumps and highly branched vegetative hyphae can be found, showing normal growth.

The test was performed on a total of three strains, just to rule out strain specificity, WT, XEG and mRFP; the protein producing strains mRFP and XEG are plasmid carrying WT-strains, thus chromosomally the same strain.

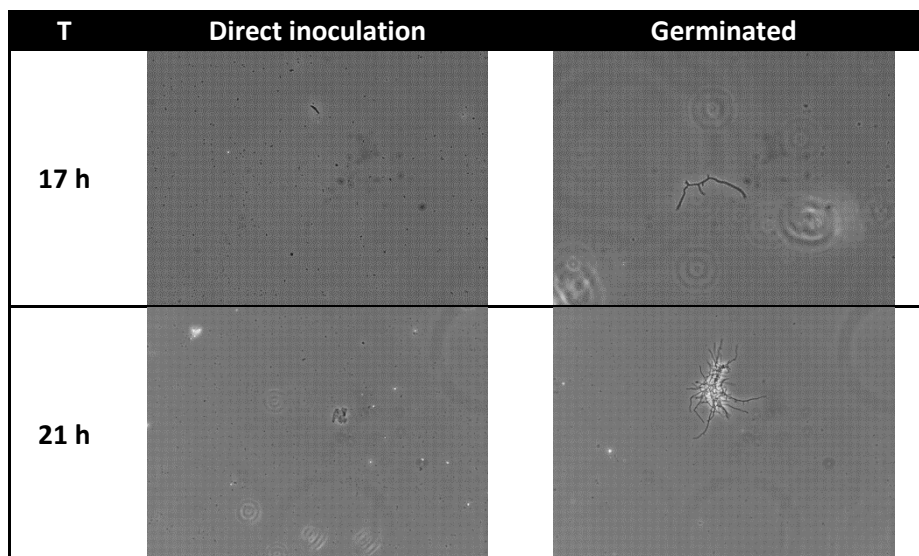


Figure 5 Microscopy pictures of the cultivation after either germination or direct inoculation, the pictures displayed are from the mRFP strain

The experiment was further monitored as intermittent measurement of OD over the following two days (for a total of 6 samples from each culture), and the highest final OD was found to be 2.02 and 10.4 respectively for the direct inoculated spores and the separately germinated spores of the mRFP strain. Despite attaining some growth, then the directly inoculated strains had a very long lag-phase, which is unfavourable, only few samples are available which made determination of growth rate difficult, though direct inoculation gave approx. $0,12h^{-1}$ and pre germination gave $0,17h^{-1}$. The mRFP strain was the only one that produced any growth in the direct inoculation. Neither the WT nor the XEG strain produced any growth without a separate germination step

Germination time

Having established that a separate germination step was required to ensure growth, a time course study was made in order to determine what duration was required to obtain the highest possible percentage of germinated spores, but the least amount of exponentially growing cells. This experiment set the precedent for how spores were handled in all subsequent experiments.

Two strains were compared: the wild type (WT) and a xyloglucanase producing strain (XEG), since these seemed to be the most susceptible to the germination conditions, in the previous experiment.

As can be seen in Figure 6, ten hours of germination, as the in-house SOP suggested for *S. coelicolor*, in a rich medium results in large multicellular aggregates, which means that a too developed stage has been reached. There is too much biomass growing already, and large visible pellets are observed. Hence, there is little hope that this biomass will form a dispersed mycelium during the $\approx 3-4$ generations that will be the exponential growth phase (10-20 hours).

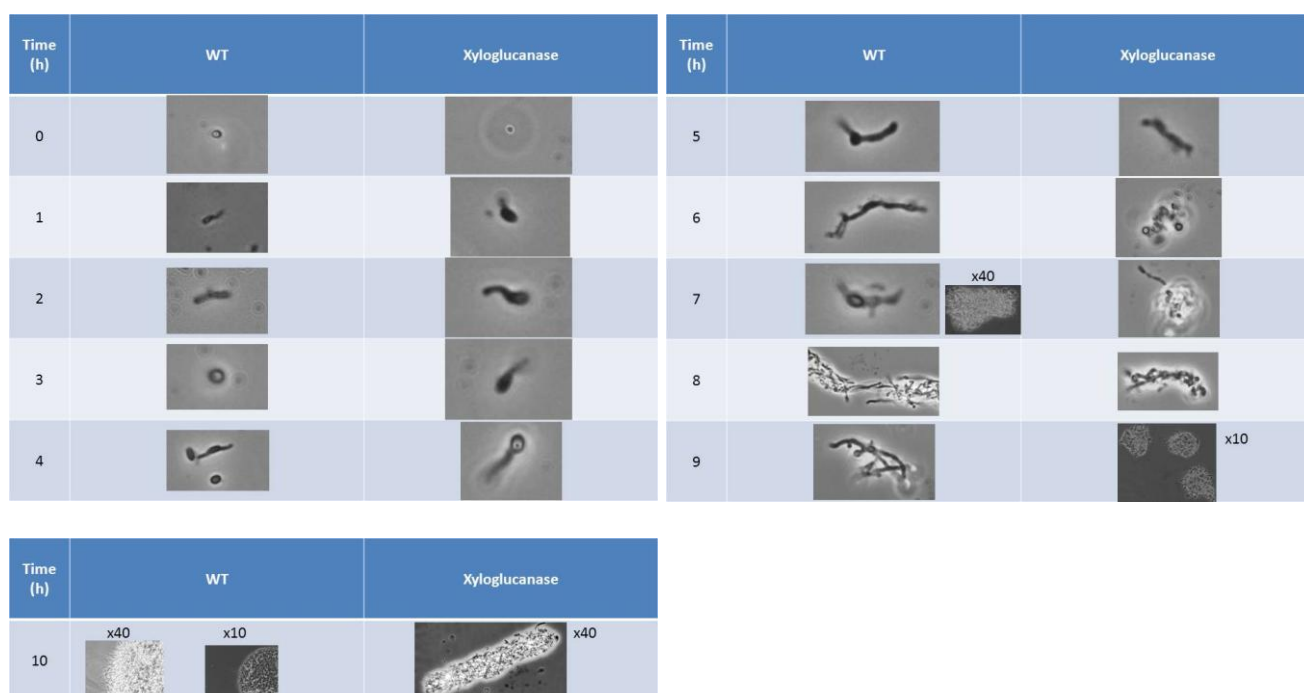


Figure 6 A time course development of germination of spores of *S. lividans* in 2YT medium. The pictures are of representative spores found in the medium.

Going back through the time points, it can be seen that the germination is asynchronous across the culture, for both strains. As can be seen for the WT at 6 and 7 hours, some spores have just started developing into vegetative hyphae ($T=7h$), whereas some multicellular aggregates are already seen ($T=7$). At $T=6h$ vegetative and branching hyphae can be found, though none of the samples screened contained large multicellular aggregates. It is likely that the large clump depicted at $T=7$ is a result of imperfect filtration of the original spore stock (see Sporestock heading in the Materials and methods section), since too many cells are present to have been formed within just 1-2 doubling times (an average growth rate between $0.1-0.2h^{-1}$ would result in doubling times of 3.5-6.9h, which makes it impossible that the large clump can be resulting from a single germinated spore, or clump of spores). No such biomass clumps were found during the first time points, but each time point represents a separate sample from the shake flask, which introduces the possibility of not detecting the exponentially growing cells at the earliest possible time.

6 hours is the latest time point where no large clumps are observed, albeit some growing and branching hyphae can be found. It is a trade-off between how long the spores require in the rich medium, before most have germinated, but not so long that they will be in exponential growth. The final decision was that 6 hours seemed to be the right balance to not be able to find still un-germinated spores, nor a large number of exponentially growing cells.

Growth experiments in batch

Having established the basics for generating reproducible growth, the production capability of various strains had to be ascertained. The wild type (WT) was used as control, both to test endogenous cellulase and xyloglucanase activity, as well as to test if the cells were encumbered by producing heterologous proteins. The producing strains investigated are mRFP, XEG and C_{elA}. All three strains expressed the product from the multicopy pIJ486 plasmid, harbouring a thiostrepton resistance marker. The mRFP strain produced a monomeric Red fluorescent protein, the XEG strain produced a xyloendoglucanase (Sianidis, et

al., 2006), and the CelA strain (CelA-19) produced a thermostable cellulase, and was provided by Olafur Friðjónsson, Matis, Iceland (Anné, et al., 2012). The initial experiments were performed in batch, and priorities were to determine the kinetic parameters: growth rate (μ_{\max}), biomass yield (Y_{sx}) and rate of product formation (potentially yield of product as well).

Initial characterisation

The growth rate can both tell whether the strain was hampered by producing the protein, thus hinting at (severe) stress, and was also a necessary parameter for designing further experiments. Future investigations were to be made in chemostats, so a plausible range of dilution rates could be determined from μ_{\max} in batch.

The batch experiments were performed in 1L lab-scale reactors, and the medium was a defined minimal medium, adapted from the medium described by Evans (Borodina, et al., 2005) (see Materials and Methods section). The reactors were inoculated with germinated spores, obtained by the method described in the previous section. Measurements were taken for analysis of extracellular metabolites (in this case substrate and product, i.e. glucose and the protein of interest, though later analysis of other metabolites would be possible), dry cell weight and OD was determined, and off gas was continually monitored for CO_2 and O_2 .

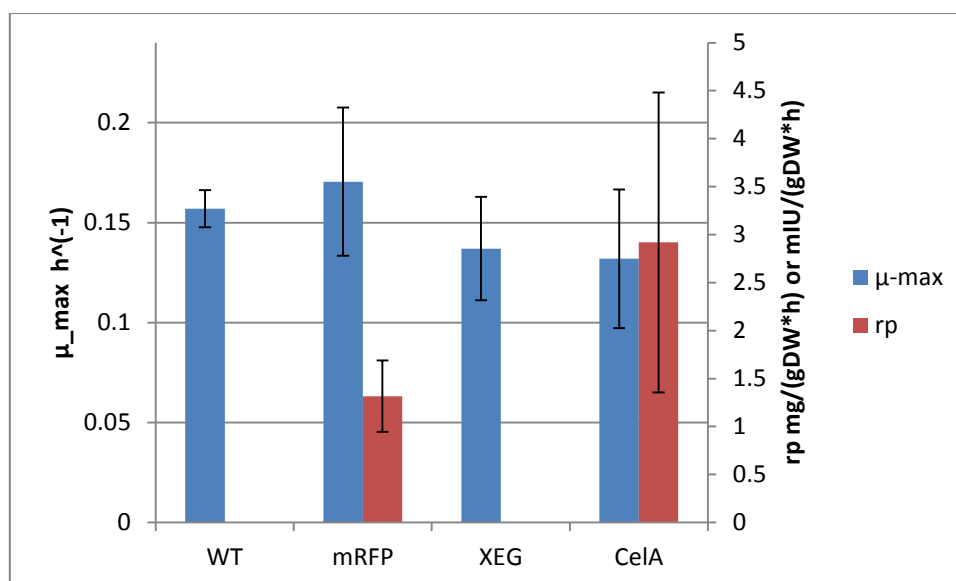


Figure 7 Growth rates determined on biomass concentration for comparing the WT and protein producing strains, μ_{\max} determined on CO_2

As can be seen in Figure 7, there was no significant difference in the maximum specific growth rate of the producing strains XEG and CelA and the non-producing wildtype strain. The mRFP strain grew significantly faster than any of the other strains, and produced a more dispersed morphology than any of the other strains, the WT included. For the wild type and the two cellulytic strains, clumps or pellets could be seen during growth in the defined minimal medium. When considering production, unexpectedly, the XEG strain did not produce any significant level of active enzyme.

Data for the XEG strain has been published (Sianidis, et al., 2006), and the strain was characterised as a high producer. The published final titer of product was 0.48 U/mL, with a maximal activity at pH 7.5 at ~50°C. In the present study AZCL-Xyloglucan from Megazyme was used for determination of xyloendoglucanase activity. No activity was observed in the fermentation samples. The assay was verified using a purchased standard enzyme (E-CELTR EGII from *Trichoderma Reesei*, from Megazyme) as positive control, and the procedure seemed to work satisfactorily in a wide concentration range of the enzyme; even when diluting the standard up to 16-20.000 fold, was it possible to get an activity reading, which in the best of cases corresponded to the endpoint readings, approximating $6.25 \cdot 10^{-3}$ U. To verify that the strain performance was not an artefact, or caused by plasmid loss in the working stock, a new vial of the strain was received from the Belgian lab (Lab of Jozef Anné, KU Leuven) that has constructed the strain and a new working stock was prepared. Yet, no discernible production was observed. The reason for this is still unknown, and it was decided to not include this strain in any of the further investigations beyond batch characterisation.

The CelA strain was assayed to give activity of the cellulase utilising carboxymethyl cellulose, as per the standard defined by (Ghose, 1987). Since it is thermostable, and had a tentative optimum temperature around 90°C, and the thermostable standard enzyme (E-CELTm, from *Thermogata maritima*, from Megazyme) was characterised at 80°C, the assay was performed at 80°C). The units are based on the amount of released sugar in a set amount of time. Since the result of the assay is an enzyme activity, the measured activity cannot directly be related to a specific amount of protein, since activities are heavily dependent on temperature, but also influenced by among others metal ions in the medium. This unfortunately makes the comparison with the mRFP strain difficult, since mRFP can be compared to a standard protein, and a concentration can be obtained. It seems the mRFP strain may have picked up a beneficial mutation since it produces a more amenable morphology, and higher growth rate. Since both of the cellulytic strains have the same growth rate, but only one of them produces any discernible amount of active enzyme, then it can be suggested that there is agreement with the finding that the growth is not affected by protein overload in the cells.

The yield of protein can only be determined for the mRFP strain, since it is the only one of the producing strains, for which the actual protein concentration in grams is available. The yield coefficient of product per substrate is 2,65 mg / g glucose, which essentially is without impact on the carbon balance.

Using an amino acid as nitrogen source

I was mostly interested in running the experiments in a defined minimal medium, since the future objectives were to run chemostats, and possibly perform a flux analysis of the producing vs. non-producing strains. A minimal medium makes the flux analysis clearer to interpret, since no alternative carbon enters the cell, and the characterisation is greatly facilitated by using a defined medium composition. A further benefit in using a defined minimal medium is the general ability to interpret results is facilitated through simplifying the conditions, and thus allows gaining a deeper mechanistic understanding of the physiology and metabolic capabilities. We had reports from a collaboration partner that it gave higher growth and product titer when using glutamate as a nitrogen source instead of ammonium. The first assumption I made upon this report was that adding glutamate instead of ammonium would also supply extra carbon (5 carbons per nitrogen); therefore the strain would essentially have the double amount of available carbon compared to our minimal medium. Though, it would be prudent to test this hypothesis, and therefore experiments were setup comparing the two media recipes side by side. To control the conditions to the

highest degree, and not depending on comparing the effects of a phosphate buffer (the MAN medium) with a MOPS buffer (DTU medium, in flasks or plates), the experiments were run in 1L lab-scale reactors at 30°C, with pH 6.8, controlled by addition of 2M NaOH or H₂SO₄, with 1VVM airflow and stirring was set to 600 RPM.

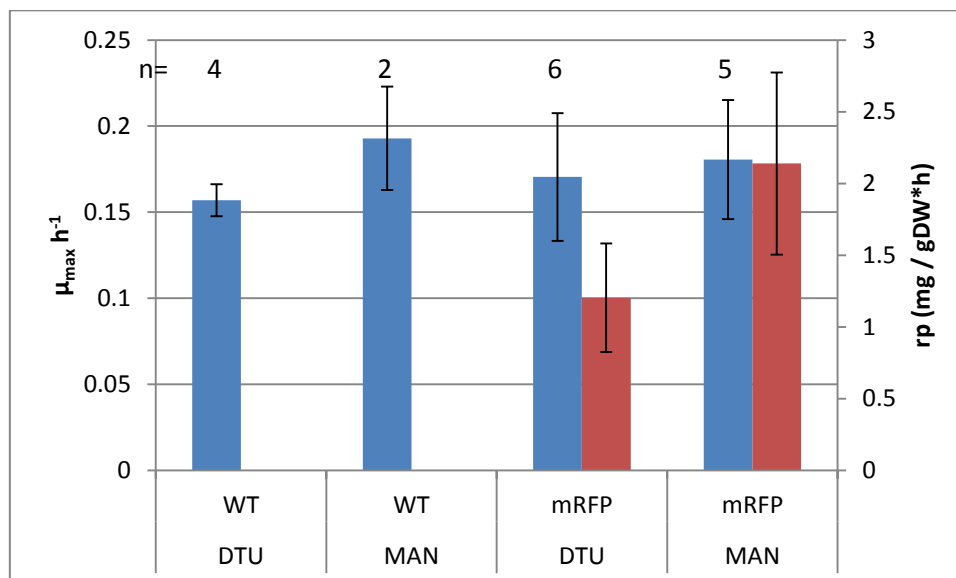


Figure 8 Estimating the effect on growth rate and product formation rate of supplementing the medium using glutamate, the growth rates are determined on CO₂ – the blue bars are growth rate, the red bars are specific production rate measured in mg/(g DW*h)

The strains are compared based on their growth rate determined from the CO₂ data, which is presented in Figure 8. When comparing the mRFP strain on the two media, the average growth rate is a little higher when using the supplemented medium compared to the standard minimal medium, though the standard deviation makes it difficult to determine whether there actually is a difference. Product formation is greatly increased in the supplemented medium. Since it is the specific product formation rate, which is normalised against biomass concentration the extra carbon can likely not be attributed to the increased rate. For the WT strain, a more marked difference in growth rate is observed, favouring the supplementation of the medium, there is obviously no product formation in the WT strain. Further, the most striking difference was the shortening of the lag-phase, when using the supplemented medium. With equal inoculum, which would normally mean that the culture entered exponential phase after 20 hours in the minimal medium, then using the glutamate supplemented medium the exponential phase was reached after about 10-15 hours. The reason why the mRFP strain seems unaffected in terms of growth rate is not clear, but since it generally produces higher growth rates and better morphology, it has likely gained a beneficial mutation during the transformation, or selection of the original transformant. The trait has been present in all stocks of the strain used in this project.

Influence of thiostrepton concentration

An experiment was made in deepwell microtiter plates, using the method described by (Sohoni, et al., 2012), where all the strains were grown with various levels of thiostrepton, to determine whether it was possible that the non-producing strain of XEG had lost the plasmid. A further verification was made in an SDS gel, to see if it was possible to detect a protein that could be inactive XEG, and to see if it was possible

to estimate the amount of CelA in grams, relative to mRFP. The levels of thiostrepton were 0 (blank) – 25 µg/mL. All strains were tested at all levels. It is mostly a qualitative test, since there is only a single end-point sample from each well of the deepwell plate.

The SDS verification was performed on the end-point samples, and no specific normalisation has been used, it is merely a screen, if it is possible to detect the cellulytic enzymes. As can be seen from Figure 9, there seems to be no product resembling XEG (~97 kDa). Despite knowing CelA is active, it is not clear which band in the lanes that could correspond to CelA (~25/28 kDa, depending on whether the *vsj* signal peptide is attached), which practically eliminates estimating the protein concentration of the CelA protein. The concentration would have been estimated by relative comparison to mRFP, of which the concentration is known from the mRFP assay. Finally the mRFP band is very visible (27 kDa); mRFP has been observed in titers up to 0.2 g/L, very close to the highest titer reported for any heterologous protein produced in *S. lividans*, though it seems that the sizes of the bands in the samples don't fit too well to the ladder. Further, it seems that a higher concentration of thiostrepton results in neither higher nor lower productivity, and there were no obvious visible effects detectable, i.e. the strains seem not to be stressed by the higher concentration. Finally, in none of the lanes does there seem to be visible signs of protease activity, as there are no smears, supporting the expectation that the lacks of activity of the XEG protein is not a result of degradation.

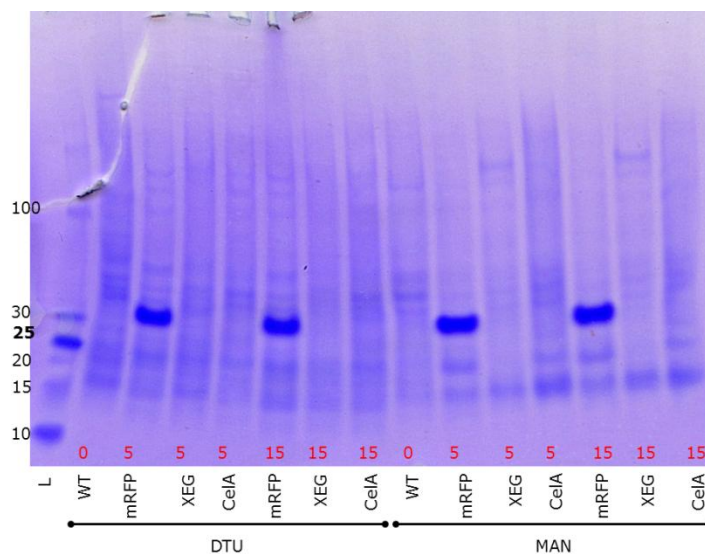


Figure 9 An SDS page gel comparing the effects of using glutamate as nitrogen source instead of ammonium, and varying levels of thiostrepton (marked in red letters in µg/mL)

Only minor differences between the two conditions were observed. It may even be suggested that there seem to be fewer extracellular proteins in the glutamate supplemented broth (MAN), as compared to the defined minimal medium (DTU), though no normalisation on protein concentration or biomass has taken place. There is a band around 100 kDa in the XEG lanes in the MAN conditions, though still no activity was measured in the assay. There is a similar, but slightly smaller, band in the WT in the glutamate supplemented medium, owing to the suggestion that no XEG is produced. The plasmid can be expected to still be present, since the selection with thiostrepton seems functional, even at high concentrations of thiostrepton.

There seem to be no inducing effect of increasing the level of thiostrepton on any of the bands, and the missing activity of the XEG protein cannot be explained, though all samples have been frozen before being assayed, possibly the XEG protein was damaged in the freeze-thawing, in a more profound way than CelsA, which retained activity, and mRFP that seemed not to be affected, since mRFP activity can be assayed, and remains relatively constant despite being subjected to several freeze-thaw cycles. Activity has not been tested before and after freezing, so some activity may have been lost.

Conclusions on the batch experiments

The protein producing strains were compared in batch, and the conclusion is that the XEG producing strain does not secrete active product, in neither the minimal nor the glutamate supplemented medium, and will therefore be left out of further studies.

The CelsA strain did produce a measurable amount of product, and suggests that *S. lividans* can be used as an expression host for thermostable proteins. The benefit of using glutamate was not characterised for the CelsA strain, though a qualitative test in an SDS gel was performed, and no clear protein band was seen, making it difficult to assign a protein concentration for the product.

The WT strain benefitted from having the medium supplemented by glutamate, in that the growth rate was improved by approx. 20%.

The mRFP strain performs very well in both bioreactor and smaller scales, where it displays an improved morphology compared to the other strains. The specific reason is unknown. Using glutamate to supplement the medium does seem to have a marked effect on production rate, and it can therefore be suggested that further investigation should be made of the effect of other carbon / nitrogen sources in the next section, and chapter 4 respectively.

Extending the range of carbon sources, functional genomics.

Although *Streptomyces lividans* has been studied as a host for producing heterologous proteins for more than 20 years (Brawner, 1994) (Anné & Mellaert, 1993) little has been reported on physiology regarding products directly derived from primary metabolism. Therefore, the full potential of *S. lividans* as cell factory has not been explored. The literature is generally focused on developing the secondary metabolites and the physiology of their production. Although some themes are commonly applicable between primary and secondary metabolites; secondary metabolites are in nature generally produced as part of an ecological adaptive strategy (Bibb, 2005), and therefore as a response to some kind of stress or limitation e.g. phosphate (Bruheim, et al., 2002).

while many various carbon and nitrogen sources have been tested, the aim has usually been to develop secondary metabolites. Though, in order to estimate the potential range of carbon sources that can be utilized as single C-sources a screen was performed using the commercial Biolog phenotype assays, plates 1 and 2 (PM01 and PM02), carbon sources (Bochner, et al., 2001). This way 190 carbon sources can easily be screened for growth and product formation. The scale of the experiment is 100µL, thus expanding the number of tested carbon sources, though limiting the amount of quantitative data from each source. It is the classical trade-of between insights vs. scope, see Figure 10.

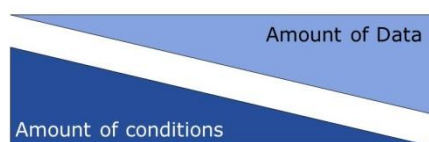


Figure 10 Showing the trade-of between the amount of data or insight that can be gained as opposed to the amount of conditions/strains that can be tested.

The screen can be used to give an initial idea of how well a carbon source can be converted to biomass and product. Any findings from the Biolog scale will have to be verified in larger scale fermentations, since the Biolog screening system among other issues relies on natural polymers to reduce evaporation, and a quantitative analysis will substantiate any finding. The system can support a functional genomics approach, and does reduce the amount of conditions that needs to be tested in a larger scale. No complex sources are included in the plates since they have been designed to give a phenotype fingerprint (Bochner, et al., 2001), and thus inform very well on the capabilities of uses of single substrates.

The experiment is handled in a liquid automation system, where frequent reading (sampling) can take place, and thus a large amount of growth and product formation data, including time-course developments, can easily be gathered.

Carbon source screening

Due to the many sample points, of the many parallel experiments, the data from the plate reader was handled in Matlab, and a simple time-course development of growth, and fluorescence respectively was produced. The raw output of the OD and RFP development over time is displayed in Figure 11 and Figure 12, to illustrate the type and amount of data that had to be handled, but also to give an impression of which carbon sources are promising. Due to the size, the data for the PM02 plate are presented in appendix A.

The axes of the plots are designed to be the same in all of the subplots, and the maximum limit on the Y-axis was defined by the plot which reached the highest value, exactly therefore the plots can be directly compared. Some of the things that can readily be observed from the growth curves in Figure 11, is that besides some of the expected good carbon sources, like amino acids, and organic acids, the 3 Tween compounds (Tween 20, 40, and 80) are readily being consumed. It seems that the longer the Tween side chains, the higher is the output from metabolising the Tween compound. This is not surprising since compounds with the molecular formulas of Tween40 ($C_{58}H_{114}O_{26}$) or Tween 80 ($C_{64}H_{124}O_{26}$), contain high amounts of carbon, and compared to e.g. glucose, then even dilute concentrations of any of the Tween compounds will support both biomass and product formation to a higher degree than a similar concentration of glucose ($C_6H_{12}O_6$). Interestingly, Tween is also used in preparation of spore stocks, which might be unfortunate, since such a rich carbon source could be expected to initiate germination. It is not something that I have been aware of before, and therefore I have not observed this, though it may very well explain the reason for keeping the spores on ice, as described by (Kieser, et al., 2000).

Getting a comprehensive overview from the time course plots is complicated, because they are intricate compared to the amount of data that is provided, therefore, in order to reduce the presentation of the data, and produce a better overview, each curve is summed up as the area under the curve (AUC), whereby each carbon source becomes represented as a single pair of values (representing amount of growth and

fluorescence), as can be seen in Figure 13. This AUC can be considered to be a measure of the Y_{sx} and Y_{sp} respectively, since it represents the total amount of biomass/product obtained from the initial substrate, as well as it indirectly contains the growth rate, since a high growth rate would yield a high AUC, and a low growth rate a low AUC. A feature that is hard to discern from the AUC is whether a value is caused by e.g. a constant mediocre growth or a high growth after a longer lag-phase, which may result in the same value. An alternative would be to take the end-points, though this method also does not inform on the growth rate, and some of the cultures seem to have a drop in the OD in the end, which would reflect negatively in their standing. Further, the signal from the RFP reading is prone to noise, because of the low amounts of protein in the small volume, but this noise is better counteracted due to the summation. Finally I chose to use the AUC, due to the reduced impact of the noise, as well as it seemed to best represent growth capacity (final attainable OD) as well as rate (to some degree, despite being susceptible to lag-phase offset).

As can be noted, the negative control in the top left corner (circled in red) has almost no growth (Figure 11) as expected, but a rather high production of RFP (Figure 12). This is both unexpected and quite unfortunate, since such a high amount of fluorescence in the negative control, would mean that production has taken place from a well that should contain no carbon. One of the problems this causes is that the obvious methods of normalising the experiment would be to subtract the negative control from all other well, after having normalised each well for their individual background measurements. First, the adjustment for the individual background readings is done by subtracting the initial measurement (time point zero) from each subsequent measurement in the related well. Hereafter the determination of the AUC takes place followed by subtracting the negative control in order to express the produced amount of biomass or product caused by the source in each respective well. This limits the reliability on the fluorescent data, and it would be advisable to verify not only all of the promising findings, but also include some of the mediocre and poor findings, to validate the experiment. The growth data seems to be unaffected by this curious result. I have had contact to other people who have expressed problems with obtaining growth in the negative control of the Biolog plates.

Further, it is not possible to use one of the wells that show no growth, instead of the negative control, because while only a very slight growth is seen in the negative control, it can be assumed that any well with no growth may be caused by a carbon source which has an inhibitory effect on the cells. And a little growth can be expected, since the cells have been germinated in a rich medium, and would therefore have a slight amount of carryover of carbon from the rich medium, thus substantiating the claim that the sources producing no growth are actually inhibitory.

The absolutely most positive aspect is to see how little the variance is between the two separate plates. In the vast majority of the wells, there is very good correlation between the replicates (The magenta and green lines represent the individual experiments, the blue line is the average between the two). There is a large number of high growth compounds, thus suggesting many useful carbon sources. The next step, where these plates support functional genomics, is to investigate which carbon sources produce any appreciable growth, and correlate these to, among others, (Hodgson, 2000), and map out active pathways. This can then be further combined with blast searching the draft genome, and thus a functional annotation can be made. The genomic and annotation work is outside of the scope of this thesis, though the results will be shortly discussed.



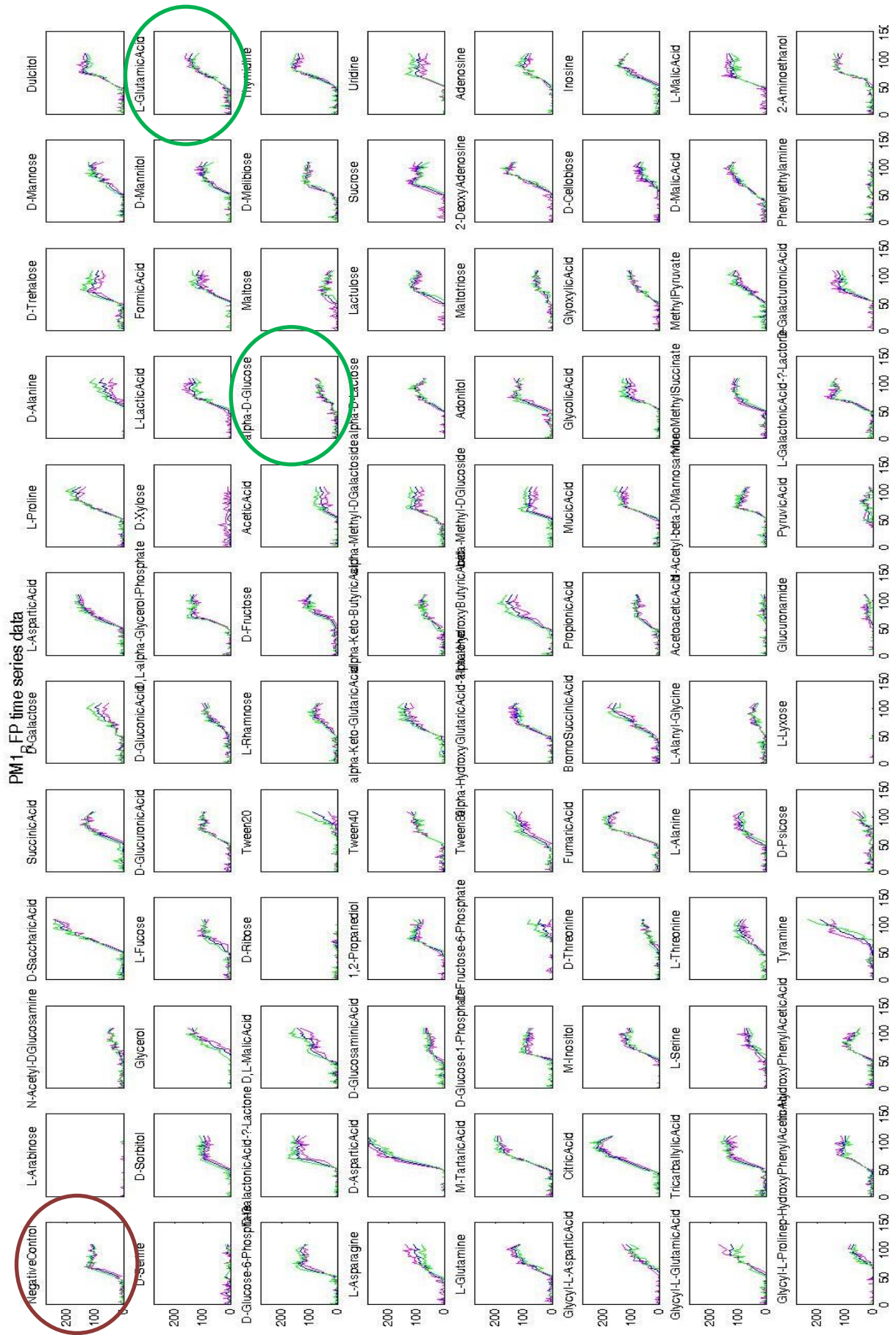


Figure 12 PM01 time course RFP vs time. The magenta and green lines represent the individual plates, and the blue line is the average between these two plates, the negative control has been marked with a red ring, and with green rings are glucose and glutamate

Unfortunately the RFP readings gave an unexpected high signal for the negative control, which means that there will be quite a number of C-sources that after calculating the AUC, and normalising the data, get a negative amount of product in Figure 13. I will accept this, while keeping in mind that a negative fluorescence is unlikely, but a result of removing the background and centring the data. It is necessary to normalise the data, i.e. divide every single AUC by the norm of the vector, because otherwise the values of the growth and fluorescence are of different orders of magnitude, and the distance measure used in the hierarchical clustering cannot take this into account. Further, the centring is required to ensure the mean=0, because the method used to determine the sources that perform remarkably well on either growth or production is finding those points that have higher values than $2 \cdot \sigma$; but this relates to the mean of the population, and therefore the values need to be higher than $\mu + 2 \cdot \sigma$, or simply centre the data, and select those points that have higher value than $2 \cdot \sigma$. Regardless, the relative positions of all points will remain the same, despite the changing of magnitude and centring of the data. Subtracting the negative control, because it is done to each point in the entire population, is just a matter of shifting the baseline, which yields some negative results for production by some carbon sources. When considering the growth only a few C-sources produce a lower growth than the negative control, which may suggest an inhibitory effect of those C-sources.

Hierarchical clustering

After having summed up the areas, a hierarchical clustering is performed in order to investigate a possible grouping of compounds, in an unsupervised manner. The hierarchical clustering is performed by using the 'Euclidean' distance measure, and in the agglomerative clustering using the 'average' metric. These are options for the built in function in Matlab, with the input being a matrix with the AUC of growth in one column, and AUC for production in the other column, together with a list of labels. First performing the clustering allowing 96 leafs, and from the dendrogram deciding that there was a natural division into 9 groups.

The scatterplots in Figure 13 displays the normalised, centred, and grouped AUC's of growth and product formation. Group 1 (red) is the group termed 'no-growth'. It is the group containing the negative control (marked with an 'x'). It seems in PM01 that there is a bit of an artefact in the hierarchical clustering methods, since the group seems to be triangular in shape, and that causes some compounds, which have higher growth than the subsequent better growing group, to be included in the "no-growth" group. A different grouping could be obtained by changing the distance measure, possibly even choosing a K-means method, though as the method is used to get a preliminary overview of the compounds, and not for thorough statistical analyses, further effort will not be made.

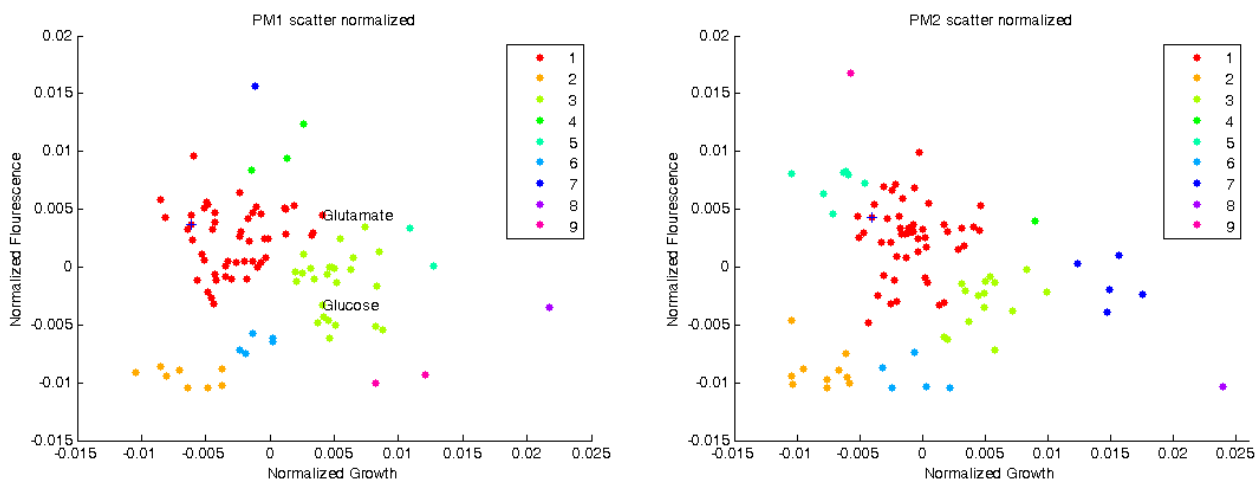


Figure 13 The scatterplots representing PM01 (left) and PM02 (right). The colours are their grouping as determined by hierarchical clustering. Each point represents 1 carbon source (including the two negative controls, further marked with an X), and is the average value of the two separate replicates.

It is interesting to note that in both plates from Figure 13 a pattern seems to suggest that when it comes to carbon sources, as opposed to nitrogen sources (in chapter 4), there is either a preference for growth OR product; no compounds are situated in the top right corner, and few are close to the axis from origo to the top right corner. This finding is somewhat surprising; since the product is expressed from a constitutive promoter thus a tight correlation between growth and production was expected. It could almost be insinuated that there is a slight tendency towards an inverted relationship between growth and product formation, as soon as some threshold is crossed.

The carbon sources across the two plates that show the highest growth (those that have higher growth values than the 2σ) are as follows: Gelatin, Laminarin, 2-Deoxy-D-Ribose, L-Histidine, Putrescine, D-Mannose, D-GlucosaminicAcid, PyruvicAcid

Gelatin is a degradation product of collagen, Laminarin is a glucan, 2-Deoxy-D-Ribose is a C5 monosaccharide, L-Histidine is an amino acid, Putrescine is a degradation product of amino acids (polyamine), D-mannose is C6 mono saccharide, D-glucosaminic acid is related to chitin degradation, Pyruvic acid is a TCA intermediate.

Laminarin has not been annotated (β -1,3- β -1,6-glucanase for coelicolor), though (Tamashiro, et al., 2012) seem to report that *S. lividans* has a xylanase with laminarin binding affinity. Chitin has been reported as substrates of *S. lividans* (Křišťfek, et al., 1999). Gelatin is hydrolysed protein, and since it can be used, it seems *S. lividans* does have extracellular proteolytic activity, which has been described for some strains, though regulation may control reduce the amount when other C-sources are present. Lysine and polyamine degradation happens via 2-ketoglutarate (Leitão, et al., 1999).

Conclusions C- sources

While the best C-sources are amino acids, and degradation products of polymers, mirroring the nature of *S. lividans* habitat it is interesting to note that none of the selected compounds are glucose, or glutamate, sources that were remarkably good in the nitrogen screen, as will be shown in chapter 4. It can be noted from Figure 13, glutamate has been grouped with the negative control, and glucose has been grouped with the slightly better growing, but not producing group (3). This is despite displaying promising growth, and for

glutamate, production in Figure 11 and Figure 12. The method is promising, but using the Biolog plates, the negative control is quite sensitive, which I have found that others have also remarked upon (Unpublished personal communications).

Summing up

One of the most significant things relevant to cultivating filamentous organisms, and specifically *S. lividans*, is that thorough attention must be paid to how inoculation takes place, to generate repeatable experiments. Many methods have been developed, and each group will have their preferred way of handling these organisms. But even if the method for inoculation is in place, the stock from which it takes place must be handled with care as well. One of the things that introduced the highest level of variability was the stock. It is difficult to say precisely what it was about the stock, but when inoculating to the same spore concentration across stocks, variability is observed in among other the growth rate. This obviously complicates matters, but it also suggests that sporulation is not completely understood, and regulation in terms of “morphological history” does have an effect in a way that is difficult to anticipate.

For the batch experiments it was found that there seem to be no adverse effect on growth rate from producing any of the 4 heterologous proteins, despite that XEG seems to not be produced in its active form. It was not possible to find the protein concentration of CelsA, though it was produced and active. mRFP is an excellent model protein, as production can be followed on-line or at-line, this model protein was found to be suited for the extended carbon screening.

For the extended carbon screening the polymers seem to be performing best in relation to growth. There was a very high, and unexpected, production of mRFP in the negative control, with very little growth, which indicates that the experiments should be performed again, to ensure better use, and more clear conclusions. In a functional genomics perspective the method is valuable, though the experiment would benefit greatly from being re-run, so that productivity can be introduced as a selective criterion.

Materials and methods

Sporestock

The spore stock was made, as described in (Kieser, et al., 2000), growing the spores for approximately 8-10 days on mannitol-soy flour agar plates. The spores are ready to be harvested when the colonies (preferably dense lawn) have changed colour from white to a darker grey, and a noticeable geosmin scent is present. The colonies do not spread very far from where they start growing, as opposed to filamentous fungi where a triple stab will cover an entire plate with mycelium, so in order to get lawn growth, a high inoculation as well as dispersing with glass beads or Drigalski spatula is necessary.

Harvesting of spores was performed according to the description of (Kieser, et al., 2000) using 0.1% Tween 80 in 15mL water, per 15cm plate). The solution is rubbed thoroughly with a Drigalski spatula, and pipetted into a 50mL Falcon tube. The solution is agitated violently to disperse the spore chains. The solution is thereafter filtered through a 5mL tip with glass wool used as filter, to remove cells. Hereafter is the solution centrifuged at 5000g for 10min at 4°C, and the supernatant discarded. The spores are resuspended with a known amount of water, depending on how many plates have been used, and an equal amount of 40% glycerol is added. The spores were stored in 210µL aliquots in 20%vol glycerol at -80°C. Of each strain a master stock has been made, which has been the progenitor of any successive working stocks.

Germination

Germination was performed in 2YT medium, the composition of which can be found in (Kieser, et al., 2000), in g/L Yeast extract 16, Tryptone 10, NaCl 5. The germination was found to occur best 6 hours at 30°C, 150 rpm in baffled shake flasks, with 30-40 4mm glass beads. After which the contents of the flask, without the glass beads, was transferred to a 50mL falcon tube and centrifuged for 10 minutes at 4000 g at room temperature, after which the supernatant was discarded, the cells re-suspended in water, and used for inoculation.

Inoculation conc.

For fermentation experiments a final concentration of spores in the liquid medium was 1×10^8 spores/L, giving an estimated start $OD_{600} \approx 0.01-0.03$. The minimal medium does not have a relevant OD absorbtion.

Medium composition

For growth in liquid culture a defined/minimal medium was used containing in g/L: $NaH_2PO_4 \cdot H_2O$ 2.76, NH_4Cl 5.35, KCl 0.75, Na_2SO_4 0.28, Citric acid $\cdot H_2O$ 0.42, $MgCl_2 \cdot 6H_2O$ 0.25, $CaCl_2 \cdot 2H_2O$ 0.18, Trace metal sol 5 mL/L, Antifoam (sigma 204) 0.2 mL/L. After autoclavation vitamin solution (1 mL/L) and separately autoclaved glucose $\cdot H_2O$ (20 g/L, giving 100 mM glucose) was added. – pH was maintained in reactor by addition of acid/base, and in the microtiter plates by addition of MOPS buffer (see subsection microtiterplates)

Trace metal solution has the following compositions in g/L: $FeCl_3 \cdot 6H_2O$ 5.406, $CuCl_2 \cdot 2H_2O$ 1.705, $ZnCl_2$ 6.815, $MnCl_2 \cdot 4H_2O$ 1.98, $Na_2MoO_4 \cdot 2H_2O$ 0.00484, $CoCl_2 \cdot 6H_2O$ 4.76, and H_3BO_3 0.618.

Vitamin solution g/L: Biotin 0.05, Ca-pantpthenate 1, nicotinic acid 1, myo-inositol 25, thiamine-HCl 1, pyridoxine-HCl 1, para-aminobenzoic acid 0.2.

Thiostrepton was added to the medium to a final concentration of 5mg/L.

The medium composition of the glutamate fortified medium (MAN) was, as received by our collaboration partner, in g/L: Glutamic acid 10, $MgSO_4 \cdot 7H_2O$ 0.6, Trace element 1mL. After autoclaving, separately add 150 mL NaH_2PO_4/K_2HPO_4 buffer (0.1M, pH 6.8) and glucose 50mL 20% (=10g/L final konc)

Trace elements fortified medium in g/L: $ZnSO_4 \cdot 7H_2O$ 1; $FeSO_4 \cdot 7H_2O$ 1; $MnCl_2 \cdot 4H_2O$ 1; $CaCl_2$ anhydrous 1.

Though vitamins were not specifically mentioned in the recipe, for comparison sake 1mL/L was added.

Bioreactors

Fermentation experiments were performed in Sartorius 1L bioreactors controlled by BioPAT towers. Temperature was 30°C, stirring 600 rpm with two 6-bladed rushton impellers, pH was kept at 6.8 ± 0.1 by addition of either 2M NaOH or 2M H_2SO_4 . Aeration was adjusted to 1VVM. Samples were taken for biomass dry weight, extra-cellular metabolites and mRFP analysis.

Microtiter plates

The microtiter experiments were based on the method described by (Sohoni, et al., 2012), with a temperature of 30°C, 150 rpm (with 50mm throw), in microtiter plates by EnzyScreen (EnzyScreen.com, Netherlands). The plates are 24 square deep well plates, with a pyramid bottom, 5mL total volume per well

and 3 mL working volume. To each well six 3mm glass beads were added to increase gas transfer and improve morphology. To control pH 100mM of MOPS was added and pH was adjusted to 6.8.

Sampling

Dry weight sampling was performed by filtrating 5 mL (3 mL or pooling of two wells for microtiter plates) culture through a 0.45µm PES membrane (Sartorius Stedim, Germany) using a vacuum pump. The supernatant was collected and stored at -18°C for later analysis of glucose and mRFP concentrations. The dry weight filters were dried in a microwave oven before and after filtration for 20 min. at 150W – allowed to cool in a dessicator, and weight was noted before and after respectively.

Concentration of glucose was analysed using a glucose dehydrogenase kit (ABX pentra Glucose HK CP, Hobira Montpellier, France). Off-gas was recorded using a Prima PRO Process Mass Spectrometer (Thermo Fisher Scientific).

mRFP concentration was obtained by measuring fluorescent activity in a Biotek plate reader, using 50µL culture, and an mCherry protein from Biovision as standard. For fluorescent measurement black-sided 96 well plates from Greiner were used. The filters used in the Biotek reader were excitation 530/25nm and emission 595/35nm

Enzymatic assays

The CMC assay is performed as a measure of endo-b-1,4-glucanase activity, according to IUPAC guidelines, as published by (Ghose, 1987). The principle of the assay is the hydrolysis of CMC by endoglucanases, and subsequent detection of the liberated glucose by reducing sugar assay.

Reagents

<p>Sodium citrate buffer 50mM Prepared from 1M Na-cit buffer</p> <ul style="list-style-type: none"> • 210 g citric acid monohydrate in 750 mL dH₂O • NaOH solids, to adjust pH to 4,3 • Adjust volume to 1 L with dH₂O • Subsequent, use 50 mL 1M Na-cit buffer, fill to 1L dH₂O and adjust to pH 4,8 <p>Note: Autoclave the 1M solution in smaller aliquots, as Na-Cit buffer is an excellent growth medium for fungi.</p>	<p>2xDNS</p> <ul style="list-style-type: none"> • 1416 mL dH₂O • 21,2 g 3,5-dinitrosalicylic acid (CAS 609-99-4) - Add and dissolve little at a time • 19,8 g NaOH solids • 306 g potassium sodium tartrate tetrahydrate (Rochelle salts CAS 6381-59-5) • 8,3 g Disodiumsulfit (Na-metabisulfit CAS 7681-57-4) • 7,6 mL Phenol (Sigma P4682-100) add in fume hood. <p>Filter solution and store at room temperature in dark bottles, as DNS reagent is light-sensitive.</p>
<p>CMC Carboxymethylcellulose (P-CMC4M. Megazyme) 2% (w/vol) in 50mM Na-Cit buffer pH 4,8</p>	<p>Glucose standard A 10 g/L Stock is prepared, from which suitable standard solutions are prepared. The standards are to be produced in the range 0,5-4 g/L</p> <p>Remember to have one blank glucose, and one blank enzyme well</p>

The assay is performed in V-shaped 96 well PCR plates

- 30 μ L CMC is added to all wells except those for the standards. (Use a multichannel pipette – have patience, the substrate is highly viscous)
- 30 μ L of enzyme (as the assay is dependent on obtaining the critical glucose release, it is suggested to make several dilutions of each sample)
- Standards: 30 μ L of glucose solutions are added to 30 μ L 50mM Na-Cit buffer pH 4,8
- The PCR plate was incubated for 30 min at assay temperature (50/90°C), using a silicone overlay, to prevent evaporation
- Place PCR plate on ice, and add 60 μ L 2xDNS reagent to **all** wells.
- Incubate 5 min at 95°C, to allow DNS reagent to react. Place on ice.
- Take 30 μ L of reaction mix and transfer to 150 μ L dH₂O in a flat bottom microtiter plate. (Mix well, both when removing sample, and when diluting)
- Read the plate at 550 nm using a microtiter plate reader.

The enzyme E-CELT_M, from *Thermogata maritimus* from Megazyme was used as thermostable standard for comparison to CelA-19.

The AZCL-xyloglucan assay is based on the AZCL-Xyloglucan substrate from Megazyme. AZCL is azurine crosslinked substrate, which releases a blue color upon digestion. The substrate is insoluble in water, but will form gel pellets. Xeg activity was measured by following the hydrolysis of AZCL-xyloglucan. 900 μ L substrate (0.2% AZCL-xyloglucan in 0.2M sodium phosphate buffer pH 6.0, 0.01% Triton-X-100) was preheated (50°C; 10 min) and then 100 μ L of pure enzyme in 0.01% Triton-X-100 or 100 μ L of sample was added. After incubation (50°C; 30 min), the reaction was stopped by transferring the samples on ice. After centrifugation (1 min; 20,000 \times g) the supernatant was transferred to a micro titer plate and was measured at 595 nm in a Biotek MX plate reader. The protocol was adapted from (Sianidis 2005), temperature was raised, as no measureable activity was found at 30°C, and the enzyme has according to the source optimum at 50-55°C. Longer incubation has also been used.

First Xeg1 xyloglucanase from *B. licheniformis* (Novozymes) was used as standard under the same reaction conditions; later E-CELT_R EGII from *Trichoderma Reesei*, from Megazyme was used.

Biolog experiments

Screening in the Biolog plates was performed with the medium composition suggested by Biolog, which is a minimal medium, at pH7, assuming the sources in the plate does not affect the pH. All important elements, except for carbon are in the medium. The carbon sources in the plates PM01 and PM02 are as defined by Biolog, see Appendix B.

The plates were incubated in a Cytomat 2 C450 (by Thermo Scientific) incubator, connected to a Hamilton (Microlab (ML) Star, with iSwap Landscape) liquid handling platform. Growth and RFP production was determined in a BioTek (Biotek SYNERGY MX "SMATBL" Monochromator based multi detection reader) plate reader with monochromator on the fluorescence part. Growth was monitored by following OD at 600nm and RFP was determined by a fluorescence reading at excitation 543nm emission 595nm with a 20nm slit-width.

The plates were incubated at 30°C and moved by the Hamilton robot to the plate reader every 2 hours for a total of 100 hours, and a final reading at 110 hours to ensure the cells had reached stationary phase.

The data was handled in Matlab, as the level of automation eases processing. 46 measurements were recorded for each of the 96 wells, of absorption at 600 nm and fluorescence of RFP respectively.

The data was arranged in matrices by well. Measurements of OD and fluorescence were placed in separate matrices. Initially each column was adjusted for any offset by subtracting the first data point. Following that, the area under the curve was estimated for each column, for OD and RFP respectively. This was to estimate the Y_{sp} and Y_{sx} , since the area under the curve signifies the accumulative amount of biomass/product obtained by the initial amount of carbon.

To investigate any grouping present, an unsupervised method was used. The hierarchical clustering was performed as an agglomerative clustering by using the Euclidean distance measure, and then the 'average' method for linking the clusters, allowing 96 leaves, thus ending with only 1 C-source per leaf.

Works Cited

- Anné, J. et al., 2012. Recombinant protein production and streptomycetes.. *J Biotechnol*, Apr, 158(4), pp. 159-167.
- Anné, J. & Mellaert, L. V., 1993. Streptomyces lividans as host for heterologous protein production.. *FEMS Microbiol Lett*, Dec, 114(2), pp. 121-128.
- Barbe, V. et al., 2011. Complete genome sequence of Streptomyces cattleya NRRL 8057, a producer of antibiotics and fluorometabolites.. *J Bacteriol*, Sep, 193(18), pp. 5055-5056.
- Bentley, S. D. et al., 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).. *Nature*, May, 417(6885), pp. 141-147.
- Bibb, M. J., 2005. Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*, 8(2), pp. 208-215.
- Bochner, B. R., Gadzinski, P. & Panomitros, E., 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function.. *Genome Res*, Jul, 11(7), pp. 1246-1255.
- Borodina, I., Krabben, P. & Nielsen, J., 2005. Genome-scale analysis of Streptomyces coelicolor A3(2) metabolism.. *Genome Res*, Jun, 15(6), pp. 820-829.
- Brawner, M. E., 1994. Advances in heterologous gene expression by Streptomyces.. *Curr Opin Biotechnol*, Oct, 5(5), pp. 475-481.
- Bruheim, P. et al., 2002. High-yield actinorhodin production in fed-batch culture by a Streptomyces lividans strain overexpressing the pathway-specific activator gene actII-ORF4.. *J Ind Microbiol Biotechnol*, Feb, 28(2), pp. 103-111.
- Celler, K., Picioreanu, C., van Loosdrecht, M. C. M. & van Wezel, G. P., 2012. Structured morphological modeling as a framework for rational strain design of Streptomyces species.. *Antonie Van Leeuwenhoek*, Oct, 102(3), pp. 409-423.
- Chater, K. F. et al., 2010. The complex extracellular biology of Streptomyces.. *FEMS Microbiol Rev*, Mar, 34(2), pp. 171-198.
- Claessen, D. et al., 2002. Two novel homologous proteins of Streptomyces coelicolor and Streptomyces lividans are involved in the formation of the rodlet layer and mediate attachment to a hydrophobic surface.. *Mol Microbiol*, Jun, 44(6), pp. 1483-1492.
- Dobson, L., 2008. *A Quantitative Study of the Relationships between Morphology, Physiology and Geldanamycin Synthesis in Submerged Cultures of Streptomyces hygroscopicus var. geldanus*, s.l.: s.n.
- Fazeli, M. R., Cove, J. H. & Baumberg, S., 1995. Physiological factors affecting streptomycin production by Streptomyces griseus ATCC 12475 in batch and continuous culture.. *FEMS Microbiol Lett*, Feb, 126(1), pp. 55-61.

- Flårdh, K. & Buttner, M. J., 2009. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium.. *Nat Rev Microbiol*, Jan, 7(1), pp. 36-49.
- Fornwald, J. A. et al., 1993. Soluble forms of the human T cell receptor CD4 are efficiently expressed by Streptomyces lividans.. *Biotechnology (N Y)*, Sep, 11(9), pp. 1031-1036.
- Ghose, T. K., 1987. MEASUREMENT OF CELLULASE ACTIVITIES. *Pure & Appl. Chem.*, Volume 59, p. 257—268.
- Grubbs, K. J. et al., 2011. Genome sequence of Streptomyces griseus strain XylebKG-1, an ambrosia beetle-associated actinomycete.. *J Bacteriol*, Jun, 193(11), pp. 2890-2891.
- Hodgson, D. A., 2000. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria.. *Adv Microb Physiol*, Volume 42, pp. 47-238.
- Hopwood, D. A., 1999. Forty years of genetics with Streptomyces: from in vivo through in vitro to in silico.. *Microbiology*, Sep, Volume 145 (Pt 9), pp. 2183-2202.
- Ives, P. R. & Bushell, M. E., 1997. Manipulation of the physiology of clavulanic acid production in Streptomyces clavuligerus.. *Microbiology*, Nov, Volume 143 (Pt 11), pp. 3573-3579.
- Kieser, T. et al., 2000. *Practical streptomyces genetics*. s.l.:John Innes Foundation.
- Kirk, S., Avignone-Rossa, C. A. & Bushell, M. E., 2000. Growthlimiting substrate affects antibiotic production and associated metabolic fluxes in Streptomyces clavuligerus. *Biotechnology letters*, Volume 22, p. 1803–1809.
- Křišťfek, V. et al., 1999. In situ monitoring of chitin degradation by Streptomyces lividans pCHIO12 within Enchytraeus crypticus (Oligochaeta) feeding on Aspergillus proliferans. *FEMS Microbiology Ecology*, 28(1), pp. 41-48.
- Leitão, A. L. et al., 1999. Inducing effect of diamines on transcription of the cephamycin C genes from the lat and pcbAB promoters in Nocardia lactamdurans.. *J Bacteriol*, Apr, 181(8), pp. 2379-2384.
- Manteca, A. et al., 2010. Quantitative proteomics analysis of Streptomyces coelicolor development demonstrates that onset of secondary metabolism coincides with hypha differentiation.. *Mol Cell Proteomics*, Jul, 9(7), pp. 1423-1436.
- Omura, S. et al., 2001. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites.. *Proc Natl Acad Sci U S A*, Oct, 98(21), pp. 12215-12220.
- Papagianni, M., 2004. Fungal morphology and metabolite production in submerged mycelial processes.. *Biotechnol Adv*, Jan, 22(3), pp. 189-259.
- Sianidis, G. et al., 2006. Functional large-scale production of a novel Jonesia sp. xyloglucanase by heterologous secretion from Streptomyces lividans.. *J Biotechnol*, Feb, 121(4), pp. 498-507.

- Sohoni, S. V., Bapat, P. M. & Lantz, A. E., 2012. Robust, small-scale cultivation platform for *Streptomyces coelicolor*.. *Microb Cell Fact*, Volume 11, p. 9.
- Tamashiro, T. et al., 2012. Critical roles of Asp270 and Trp273 in the $\hat{1}\pm$ -repeat of the carbohydrate-binding module of endo-1,3- $\hat{2}$ -glucanase for laminarin-binding avidity. *Glycoconjugate Journal*, 29(1), pp. 77-85.
- van Wezel, G. P. et al., 2006. Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering.. *Appl Environ Microbiol*, Aug, 72(8), pp. 5283-5288.
- Ventura, M. et al., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum.. *Microbiol Mol Biol Rev*, Sep, 71(3), pp. 495-548.
- Vrancken, K., Mellaert, L. V. & Anné, J., 2010. Cloning and expression vectors for a Gram-positive host, *Streptomyces lividans*.. *Methods Mol Biol*, Volume 668, pp. 97-107.
- Wang, L., Ridgway, D., Gu, T. & Moo-Young, M., 2005. Bioprocessing strategies to improve heterologous protein production in filamentous fungal fermentations.. *Biotechnol Adv*, Mar, 23(2), pp. 115-129.
- Whitaker, A., 1991. Actinomycetes in Submerged Culture. *Applied Biochemistry and Biotechnology*, January, 32(1-3), pp. 23-35.
- Wohlleben, W., Stegmann, E. & Süssmuth, R. D., 2009. Chapter 18. Molecular genetic approaches to analyze glycopeptide biosynthesis.. *Methods Enzymol*, Volume 458, pp. 459-486.

Chapter 4. Screening and upscale

The content of this chapter is the paper titled:

“Rational Selection of Nitrogen Source for Optimized Heterologous Protein Production Using Streptomyces lividans”

Submitted to Applied Microbiology and Biotechnology on 28-05-2013

Rational Selection of Nitrogen Source for Optimized Heterologous Protein Production using *Streptomyces lividans*

Stig Rattleff, DTU-Systems Biology

Jette Thykaer, DTU-Systems Biology

Anna Eliasson Lantz, DTU-Systems Biology

ael@bio.dtu.dk Phone: +45 45252680, Fax: +45 4588 4148

Keywords: streptomyces, lividans, physiology, heterologous protein, nitrogen source, minimal medium, screening

Abstract

The medium composition has great effect on the production of heterologous protein in bioreactors. It is a parameter that can be adjusted regardless of GMO concerns or knowledge of genomic sequence.

Actinomycetes are widely known for production of antibiotics, though as hosts for heterologous protein expression they show great potential which should be further developed. *Streptomyces lividans* is especially interesting due to low endogenous protease activity and the capability to secrete proteins to the medium. As saprophyte it also has the ability to use a diverse range of substrates including cellulose. Furthermore, a growing array of genetic tools has been developed, while annotation is expected to follow.

Having used a strain of *Streptomyces lividans* that produced mRFP as a model protein a potential high throughput method was tested for optimizing medium composition. Focus was placed on selecting a nitrogen source to improve heterologous protein production. A large number of nitrogen sources were tested. The method had the great advantage that the initial steps had a high degree of automation, which allowed retaining a relatively high number of candidates. Further a substantial deal of physiological knowledge was gained about the *S. lividans* model strain with no published sequence.

From 95 potential nitrogen sources, the seven best were selected for testing in milliliter scale, of these the 3 best were chosen to be tested in 1L scale. The best N-sources seemed to retain their mutual performance across the scales, indicating the validity of this approach.

Introduction

Microbial cells have long been used as production hosts for both metabolites and proteins. Using microbes as production hosts is useful because of their high growth rate and, depending on the choice of host, straightforward biology (Zhang et al., 2010). Many compounds can be produced from an endogenous host, but typical limitations for production are stringent regulation or competing by-products. The former can result in low productivity, the latter in low yield and complicated purification of the wanted product. The possibilities of optimization are related to either genetic engineering, should knowledge of the genomic sequence be available, or else classical strain improvement and medium optimization schemes.

For protein production, as for other biological products, using endogenous hosts tends to result in low product yields. With heterologous hosts, the challenge lies in avoiding problems with folding, secretion, and product degradation in the extracellular medium. These are all problems to be addressed in each specific case, and have received much attention in literature (see e.g. (Mahmoud, 2007; Mattanovich, et al., 2012; Punt, et al., 2002)). There are many available hosts for protein production, spanning all kingdoms of life. For bacteria, the most well-known are *Escherichia coli* and *Bacillus subtilis*. Among yeasts we find *Saccharomyces cerevisiae* and *Picchia pastoris* and among filamentous fungi, *Aspergillus spp.* and *Trichoderma spp.* are well known industrial protein producers (Van Ooyen, et al., 2006; Morello, et al., 2008; Westers, et al., 2004; Van den Homberg, et al. 1997; Macauley-Patrick, et al., 2005; Keränen & Penttilä, 1995). A parameter of high importance is product stability. It is widely recognized that proteases are of detriment to heterologous protein production. Typically filamentous fungi and bacilli with high specific productivities suffer from also having high protease production (Morello, et al. 2008; Vitikainen, et al. 2005; Van den Homberg, et al., 1997). Many solutions have been attempted, among others deletion of genes encoding extracellular proteases, with varying degrees of success. Even strains with several genes deleted tend to still have remaining protease activity (Punt, et al., 2008).

Streptomyces lividans is a filamentous gram positive bacterium belonging to the order *Actinomycetales*. The actinomycetes are recognized to produce 60% of the known antibiotics (Kieser, et al., 2000). In this order several enzymes have been produced in large scale and thus can be considered industrially well known (Binnie, et al., 1997; van Wezel, et al., 2006; Brawner, 1994). Though, more knowledge is still needed before they can be considered general hosts. Their most prominent features are their excellent capability of secreting proteins to the medium and, especially for *Streptomyces lividans*, the low endogenous protease activity (Jain, et al., 2009; Vrancken & Anné, 2009; Strickler, et al., 1992).

Whether using an endogenous or an engineered host; the strain will react to the environment experienced during the process, which will impact the outcome. Process parameters that may influence the productivity and yield include pH, aeration, agitation/stirring, temperature, and substrate availability. Optimizing these parameters can be very time consuming, since it is a multidimensional problem, with many of the parameters being to some degree interdependent. The classical approaches in process optimization regard media composition and feeding strategies. When addressing media composition, several aspects need to be considered. Complex carbon/nitrogen sources are cheaper, but result in higher batch-to-batch variation, as well as set higher demands on the product purification. If the product is therapeutic then regulatory affairs may become simpler using defined, salt-based media, and downstream processing will be eased. Further, depending on the product, the choice of a carbon/nitrogen source that best supply precursors for the product can be beneficial. For the production of heterologous proteins, the nitrogen source is particularly interesting (D'Huys, et al., 2011; Voelker & Altaba, 2011; Macauley-Patrick, et al., 2005; Hodgson, 2000).

We wished to investigate the effect of various nitrogen sources on growth and protein production in *S. lividans* and to find a systematic way of testing a large range of nitrogen sources efficiently. Hence, a fluorescent protein (mRFP) was chosen as a model protein as it can be easily detected, also on-line. Furthermore, we have taken advantage of the commercial BioLog system, a system that has been designed to phenotypically identify microbes based on their behaviour when subjected to various sources of carbon, nitrogen, phosphorous, sulphur, pH and osmolarity. One of the plates of the phenotype assays contains 95

different nitrogen sources (Bochner, et al., 2001), which allowed an initial screen, with a minimum input of manual labour. Hereby, both growth and product formation were detected over the course of the experiment (in 100 μ L scale) and then used to estimate the performance of the various compounds. Based on these estimations a subset of nitrogen sources were selected and experiments repeated in 3mL liquid microtiterplate cultures. The larger scale enables a more precise estimation of product formation, determination of possible by products, and determination of the yield of product on substrate. A much higher degree of physiological data can be gathered, while keeping the amount of manual labour per cultivation to a minimum. Moreover, improved mass transfer can also be achieved compared to the 100 μ L scale, resulting in better performance and reproducibility. Lastly, a smaller selection of N-sources was selected for verification in bio-reactor scale, i.e. 500-1000mL, where highly controlled conditions and on-line off-gas analysis enables statistical verification of kinetic parameters and yields. The increased volume available also allows for an expanded set of analyses and hence, provides further physiological insight into the effect of different nitrogen sources on metabolism.

Materials and Methods

Strains and media

The strain was *Streptomyces lividans* TK24, expressing mRFP from pIJ486 (Kieser, et al., 2000) plasmid bearing a thiostrepton resistance marker, as well as the sequence for mRFP behind the strong *vsj* promoter and signal peptide (Sianidis, et al. 2006; Pozidis, et al., 2001; Lammertyn, et al., 1997). The strain was kindly provided by the lab of Prof. Jozef Anné, Catholic University Leuven. The medium used for germination was 2YT containing in g/L: Yeast extract 16, Tryptone 10 and NaCl 5 (Kieser, et al., 2000). Solid sporulation medium used was MS agar containing in g/L: Mannitol 20, Soy flour 20, agar 20. Spore stocks were prepared according to (Kieser, et al., 2000), to give a final concentration of glycerol in the stock of 20%vol and spores were stored at -80 C until use.

For growth in liquid culture a defined/minimal medium was used containing in g/L: NaH₂PO₄·H₂O 2.76, N-source supplying 100mM nitrogen (weight according to table 1), KCl 0.75, Na₂SO₄ 0.28, Citric acid·H₂O 0.42, MgCl₂·6H₂O 0.25, CaCl₂·2H₂O 0.18, Trace metal sol 5 mL/L, Antifoam (sigma 204) 0.2 mL/L. After autoclavation vitamin solution (1 mL/L) and separately autoclaved glucose·H₂O (20 g/L, giving 100 mM glucose) was added.

Trace metal solution has the following compositions in g/L: FeCl₃·6H₂O 5.406, CuCl₂·2H₂O 1.705, ZnCl₂ 6.815, MnCl₂·4H₂O 1.98, Na₂MoO₄·2H₂O 0.00484, CoCl₂·6H₂O 4.76, and H₃BO₃ 0.618.

Vitamin solution g/L: Biotin 0.05, Ca-pantpthenate 1, nicotinic acid 1, myo-inositol 25, thiamine-HCl 1, pyridoxine-HCl 1, para-aminobenzoic acid 0.2.

Thiostreptone was added to the medium to a final concentration of 5mg/L.

Biolog experiments

The Biolog experiments were performed using the medium composition recommended by Biolog for *Streptomyces spp.* The working volume in the Biolog plates is 100 μ L in a semisolid medium containing a gelling agent to minimize evaporation. This experiment was based on the PM03 microbial phenotype plate, containing 95 different N-sources.

The plates were incubated in a Cytomat 2 C450 (by Thermo scientific) incubator, connected to a Hamilton (Microlab (ML) Star, with iSwap Landscape) liquid handling platform. Growth and RFP production was determined in a BioTek (Biotek SYNERGY MX "SMATBL" Monochromator based multi detection reader) plate reader with monochromator on the fluorescence part. Growth was monitored by following OD at 600nm and RFP was determined by a fluorescence reading at excitation 543nm emission 595nm with a 20nm slit-width.

The plates were incubated at 30°C and moved by the Hamilton robot to the plate reader every 2 hours for a total of 100 hours, and a final reading at 110 hours to ensure the cells had reached stationary phase.

Microtiter experiments

The microtiter experiments were based on the method described by (Sohoni, et al., 2011), with a temperature of 30°C, 150 rpm (with 50mm throw), in microtiter plates by EnzyScreen (EnzyScreen.com, Netherlands). The plates are 24 square deep well plates, with a pyramid bottom, 5mL total volume per well and 3 mL working volume. To each well six 3mm glass beads were added to increase gas transfer and improve morphology. To control pH 100mM of MOPS was added and pH was adjusted to 6.8.

The culture was germinated in rich 2YT medium for 6 hours, in baffled shake flasks containing 20-30 4mm glass beads. The germination medium was inoculated with approx. 1×10^8 spores. At the end of germination phase, the liquid culture was centrifuged at 4000g for 10 min, the supernatant was discarded, the pellet was washed, centrifuged again, supernatant discarded, and the pellet resuspended in 5mL sterile water.

A master medium was prepared containing everything but the nitrogen source. 90 mL of the master medium was transferred to 7 sterile flasks, to which 10 mL sterile nitrogen source was added to a concentration according to table 1. During the cultivation samples were taken at regular intervals. With each sample an entire well was harvested, and for the low biomass concentrations, two wells were pooled to obtain sufficient amount of biomass for dry weight analysis.

Fermentation experiments

Fermentation experiments were performed in Sartorius 1L bioreactors controlled by BioPAT towers. Temperature was 30°C, stirring 600 rpm with two 6-bladed rushton impellers, pH was kept at 6.8 ± 0.1 by addition of either 2M NaOH or 2M H₂SO₄. Aeration was adjusted to 1VVM. Samples were taken for biomass dry weight, extra-cellular metabolites and mRFP analysis.

Dry weight sampling was performed by filtrating 5 mL culture through a 0.45µm PES membrane (Sartorius Stedim, Germany) using a vacuum pump. The supernatant was collected and stored at -18°C for later analysis of glucose and mRFP concentrations. The dry weight filters were dried in a microwave oven before and after filtration for 20 min. at 150W – allowed to cool in a dessicator, and weight was noted before and after respectively.

Concentration of glucose was analysed using a glucose dehydrogenase kit (ABX pentra Glucose HK CP, Hobira Montpellier, France). Off-gas was recorded using a Prima PRO Process Mass Spectrometer (Thermo Fisher Scientific).

mRFP concentration was obtained by measuring fluorescent activity in a Biotek plate reader, using 50µL culture, and an mCherry protein from Biovision as standard. For fluorescent measurement black-sided 96 well plates from Greiner were used. The filters used in the Biotek reader were excitation 530/25nm and emission 595/35nm

Yield coefficients were calculated as the linear correlation between two components, in the exponential growth phase. The yield of product with regard to biomass, Y_{xp} , is defined as the specific rate of product formation divided by the specific rate of biomass formation, which is the same as μ_{max} during the exponential growth phase. This means that the specific rate of product formation (rp) can be calculated as $Y_{xp} \text{ (g/gDW)} * \mu_{max} \text{ (h}^{-1}\text{)} = rp \text{ in g/(gDW*h)}$.

Data handling

The data was handled in Matlab, as the level of automation eases processing. 46 measurements were recorded for each of the 96 wells, of absorption at 600 nm and fluorescence of RFP respectively.

The data was arranged in matrices by well. Measurements of OD and fluorescence were placed in separate matrices. Initially each column was adjusted for any offset by subtracting the first data point. Following that, the area under the curve was estimated for each column, for OD and RFP respectively. This was to estimate the Y_{sp} and Y_{sx} , since the area under the curve signifies the accumulative amount of biomass/product obtained by the initial amount of carbon.

To investigate any grouping present, an unsupervised method was used. The hierarchical clustering was performed as an agglomerative clustering by using the Euclidean distance measure, and then the 'average' method for linking the clusters, allowing 96 leaves, thus ending with only 1 N-source per leaf.

Results

Initial Screening of 95 nitrogen sources

Working in small volumes in an automated environment allows handling of many parallel experiments. Therefore, to facilitate screening of a large variety of nitrogen sources, a *Streptomyces lividans* strain expressing mRFP was cultivated on plates from the Biolog system and analysed using a standard plate reader. Biolog plates have been designed to give a phenotypical fingerprint, and thus span a wide range of nitrogen sources (95 different), selected to give optimal information on the biology of the organism (Bochner, et al., 2001). Using a standard plate reader coupled to a robotic system instead of the Omnilog system, delivered by Biolog, has the distinct advantage of being able to read not only optical density (OD), but also fluorescence (RFP), besides the automated handling of the plates.

Development of OD and RFP over time was followed and used for selecting a subset of promising nitrogen sources. The criteria for selecting candidates were a high final product titer in addition to high growth. It may be argued that growth and production rates would be preferred selection criteria, though due to the RFP data being noisy at this scale, it was decided to first calculate rates in the subsequent, larger-scale steps where more sample volume is available. To reduce the complexity of the data, as well as to minimize the influence of the noise in the fluorescent data, in this first step all growth and production time course plots were converted to the area under the curve (AUC), resulting in 2 data points per nitrogen source. In order

to investigate a possible relation between nitrogen sources, the data was grouped using a hierarchical clustering routine, and then manually curated. The hierarchical clustering algorithm identified 8 clusters, based on the performance in relation to growth and product formation. The nitrogen sources included in each cluster are listed in table S1 (in supplementary materials).

For further evaluation, the data was plotted as growth vs. production and the data points coloured according to the group association obtained by the hierarchical clustering (figure 1 –scatterplot-). As assumed a vast number of nitrogen sources performed unfavourably, as can be seen from the large number of data points grouping around the negative control, group 8 (light purple). For some nitrogen sources less growth than for the negative control was observed, indicating that these nitrogen sources are inhibitory. Nitrogen sources belonging to the groups 4 (dark green), 5 (light blue) and 6 (dark blue) favoured growth, at the expense of product formation. These groups displayed a fluorescence similar to that seen for nitrogen sources grouping with the negative control and therefore will not be regarded as possible candidates for further investigation. Likewise, some nitrogen sources seemed, despite disfavouring growth, to support high levels of fluorescence; one nitrogen source in this group (group 7, dark purple) showed less growth than the negative control and the other had a small amount of growth, though not more than that of group 8 (light purple), containing the negative control. The cause of the high fluorescence is unclear; however, it seems unlikely that high amounts of mRFP were produced by the low amount of biomass observed. Data for all wells were normalised according to the first measurement in the respective well, therefore the fluorescence should not be contributed to by the nitrogen source.

It can be assumed that beneficial nitrogen sources would be found in the scatter plot (figure 1) along or above a diagonal line going through origo, but separating from group 8, containing the negative control. Using this criterion, we selected glutamate, aspartate, Ala-Gly peptide, allantoin, D-glutamic acid, histidine, L-pyroglutamic acid and nitrate as being the most promising nitrogen sources for further investigation. The chosen nitrogen sources belong to the groups 1 (red) + 2 (yellow) + 3 (light green) from the hierarchical clustering. As an alternative method to decide which nitrogen sources were interesting candidates, it was tested which nitrogen sources lie outside of the 95% region of the normal distribution. This was done testing whether each point was more than 2 times the standard deviation ($2 \cdot \sigma$) away from the mean. The test was independently performed on growth and fluorescence, respectively. The analysis summarised in figure 2 showed that 33 nitrogen sources were selected using growth as parameter; 10 nitrogen sources distinguished themselves on production. The 7 nitrogen sources chosen were those that distinguished themselves on both parameters, and the result of the test confirmed that nitrogen sources from groups 1+2+3 were those that behaved significantly different from the average regarding growth and fluorescence and hence were interesting candidates for further investigation.

Verification of physiological performance on selected nitrogen sources in 3-mL scale

In order to gain increased physiological insight, as well as to verify the findings of the Biolog experiments, 3 mL-scale submerged cultivations with 6 of the most promising nitrogen sources were performed, these were as mentioned above glutamate, aspartate, allantoin, histidine, L-pyroglutamic acid and nitrate. In addition, ammonium was included as it is the most commonly used nitrogen source and hence served as a reference. D-glutamate was deselected due to price, and the Ala-Gly peptide was considered too specific to be included.

This 3 mL scale enables biomass dry weight measurement, and thereby a more accurate calculation of the yield of g product per g DW of biomass (Y_{xp}). Y_{xp} is a value that enables comparison between different growth conditions because the rates have already been taken into account due to the definition of the yield, as described in materials and methods section. Further, with a larger sample volume available, the noise in the fluorescence data was greatly reduced, albeit the sampling frequency was lowered. All parameters were normalized to biomass concentration, ensuring optimal comparability between experiments. For medium composition, we have chosen a standard defined medium for actinomycetes, as described by (Borodina, et al. 2005). Further, to increase comparison between the conditions, we decided to keep the absolute level of glucose constant, as well as to normalize the nitrogen source to 100mN-mol/L. The total amount of carbon will hence vary between experiments, but this was considered to be more comparable, than varying the amount of nitrogen or glucose to maintain a fixed C:N ratio.

Out of the 7 nitrogen sources tested, pyroglutamic acid and histidine did not support growth to a level where a growth rate could be calculated. Growth rates observed for the remaining 5 nitrogen sources ranged from 0.07 to 0.13 h⁻¹, with the highest one estimated for glutamate closely followed by aspartate (figure -3 growth/product plot-). Ammonium and nitrate were almost indistinguishable, and the lowest growth rate was obtained for allantoin, around 60% of that observed for ammonium.

In terms of productivity, the highest values were obtained for cultivations with glutamate or aspartate as nitrogen source, having specific productivities 4 times the level seen for cultivations with ammonium (figure 3). Nitrate and ammonium performed almost equally well and lastly for cultivations with allantoin the specific productivity was less than 50% of the productivity observed when ammonium was used as nitrogen source.

As can also be seen in figure 3 the two predominant sources were still glutamate and aspartate in terms of both productivity and growth. The differences in growth rate between the amino acids and the inorganic N-sources were less than 20% in favor of the amino acids, though there was a large difference in specific product formation rate in favor of the amino acids.

The results when using ammonium as nitrogen source were similar to what we have experienced earlier. The same was true for the faster glutamate. Why ammonium was deselected in the automatic screen is unclear, since ammonium and nitrate seem to result in the same growth rate, and production (data not shown).

In this milliliter scale liquid culture the undoubtedly best results were obtained using aspartate or glutamate as nitrogen sources. The inorganic sources were in the second best group, and allantoin is the only nitrogen source that can be considered 'exotic' that produced any appreciable growth, and at least in the screening produced enough product to be selected.

Detailed characterisation of growth and production on selected nitrogen sources

To further investigate the findings, and gain a more comprehensive physiological insight, the 4 nitrogen sources aspartate, allantoin, nitrate, and ammonium (reference) were chosen for further characterization in bench-scale cultivations, glutamate had been left out since there was no relevant difference between glutamate and aspartate. In the bioreactors, the experimental conditions can be controlled to a much greater degree than what is possible in microtiter plates. The most crucial advantage is the on-line control of pH by addition of acid or base, as opposed to relying on a buffering system in the smaller scales. This is especially important in the comparison of the nitrogen sources, since it eliminates any concern of a pH related effect. Precise control of aeration, temperature and agitation is of further merit to the bioreactors, as well as the recording of off-gas data and dissolved oxygen.

Typical cultivation profiles are presented in figure 4 showing the development of biomass, product formation, glucose consumption and accumulated CO₂ over time for the selected nitrogen sources. From both the CO₂ evolution rate and the biomass accumulation data it can be seen that with all nitrogen sources 2 growth phases occur. For ammonium and aspartate, the two experiments with the highest growth rates, there was a marked reduction in growth rate in the second growth phase caused by oxygen limitation, as verified by data from the DO-probe (data not shown). There was still carbon left, and a further increase in both biomass and product was seen. For ammonium and nitrate it is important to note that there is only one C-source whereas allantoin and aspartate both supply extra carbon besides nitrogen. This could be seen, as allantoin had 2/3's of its glucose left, when a limitation, likely nitrogen, set in. For allantoin the growth was not as affected when the CO₂ curve changed, but growth stopped, when glucose ran out, simultaneous with the onset of oxygen limitation (data not shown).

Product formation is, because of the constitutive promoter used for mRFP expression, likely to be directly dependent on growth rate. As could be seen from the product profiles, product was produced simultaneously during growth. In the experiments using ammonium as N-source, glucose was not depleted until after the onset of oxygen limitation, and product was still formed though at a slower rate; just as for biomass formation. Even after depletion of glucose there was some growth, suggesting further degradation of primary metabolites. The same picture was seen for aspartate, though glucose seemed to be used after aspartate, since the reduction in concentration was offset in time, when compared to ammonium and nitrate

As can be seen from the filled symbols in figure 3, the compounds sorted by growth from highest to lowest were aspartate, ammonium, allantoin and nitrate. The growth rates spanned the range: 0,11 – 0,19 h⁻¹. Previous experiments in-house, but with a different medium composition, places glutamate in the same scale – around 0,2 h⁻¹ (data not shown). For the bioreactor experiments nitrate as nitrogen source was located near allantoin, at 70% relative to ammonium.

In terms of productivity the same picture was seen – aspartate as nitrogen source resulted in the highest production, followed by ammonium, allantoin and lastly nitrate. Aspartate has a 4 times higher specific product formation rate, relative to ammonium, and allantoin in terms of productivity performs similar to ammonium. Nitrate gave the lowest product formation rate.

Using allantoin as nitrogen source resulted in higher growth and product formation rates in the bioreactor experiments compared to the microtiter plate experiments; whereas nitrate as nitrogen source has

deteriorated on both accounts in the larger scale. In the bioreactors both productivity and growth were lower when using nitrate rather than allantoin as nitrogen source, the reason is unclear, but pH could be a factor. Aspartate kept its relative position to ammonium, and remained the fastest growing and highest producing of the tested nitrogen sources.

This would suggest that aspartate was used as both nitrogen and carbon source in the beginning of the fermentation, and later glucose became the carbon source. Alternatively they could be co-metabolized, with glucose supplying glycolysis, and aspartate supplying TCA. For allantoin the reduction in production rate coincided with depletion of glucose, and oxygen limitation. Over all, nitrate had a longer lag-phase, and slower growth rate, therefore product titers were still so low, that the estimation of a yield was complicated.

It was seen that ammonium, allantoin and aspartate retained their mutual positions/performances, across the three scales. Why nitrate changed so drastically between the milliliter and liter scale is yet unknown.

Discussion

We have successfully demonstrated a method for rational selection of nitrogen source. As an initial screen the Biolog system is extremely useful, since it covers a wide array of nitrogen sources. A large dataset providing a great deal of physiological information, despite the minimal input of labour, is generated. Though, at such a small scale there is some uncertainty tied to the results, and for a production study the actual biomass concentration should be known. Therefore, as a basis for selecting a suitable nitrogen source for heterologous protein production, further experiments, using a subset of nitrogen sources, have to be performed in a larger scale. We chose to use a two-step procedure applying a 3mL and a 1L scale. The 3 mL scale gives the required sample volume for making dry weight assessment, and thus it is possible to determine the specific growth rate, and the Y_{xp} , from which the specific production rate can be deduced. In the larger scales we define the medium composition, and can therefore choose the exact premises for the comparison. For the 1L scale, not only is there more sample volume available, but the amount of data that can be collected also increases significantly. Depending on the probes available, on-line monitoring of pH, dissolved oxygen, redox potential and optical density is possible, likewise is the recording of the off-gas that allows measuring CO_2 , O_2 , and other volatiles, such as ethanol and methanol.

All of the nitrogen sources in the screen were defined. In terms of industrial relevance, it would have been prudent to add some complex nitrogen sources, such as corn steep liquor. However, an important aim of the present study was to gain a physiological understanding of the impact of the nitrogen source on protein production. The fact that it is very challenging, if not impossible, to obtain a detailed, quantitative understanding of metabolism when using complex substrates, led us to leave these types of substrates out in favor of interpretability of the data.

Nitrogen source groups

Examining the scatterplot (Figure 1), the 7 selected compounds are located in the top right part of the plot. As stated in the results section there are a total of 33 sources selected on growth alone, though many of these compounds did not show an appreciable amount of production. This is very interesting, since the product is expressed from a constitutive promoter from a plasmid, and as such a close growth-production relationship would be expected. Why this is not the case is so far unknown. The findings of the selection

process are summarized as a Venn diagram in figure 2. The groups, as they are found in figure 2 will be discussed in the following sections, as will the observations presented in the Venn diagram

High performing nitrogen sources

The amino acids that produced the highest amount of growth and product, across all three scales were glutamate, and aspartate. These two amino acids performed remarkably well compared, not only to other amino acids but also, to the entire group of nitrogen sources. Why are these two so efficient? The reason may be that they enter directly into the TCA cycle when being metabolized. Aspartate is converted to oxaloacetate and glutamate via a transamination step; glutamate is together with NAD(P)^+ converted to α -ketoglutarate + $\text{NAD(P)H} + \text{NH}_3 + \text{H}^+$ (Hodgson, 2000; Voelker & Altaba, 2001; D’Huys, et al., 2012). A follow up question is then why Alanine was not represented among the high performing nitrogen sources. Alanine likewise enters directly into TCA, but at the pyruvate node. Alanine is formed by condensation of pyruvate with ammonium which uses reducing power, the catabolism is via the same reactions though in reverse; the responsible enzyme is alanine dehydrogenase (ADH) (Hodgson, 2000). Alanine enters TCA at the pyruvate node, and was expected to perform better – though the cost of reduction power by the ADH enzyme could be speculated to be the cause of the lower performance. It may therefore be either transport related, or it could be hypothesized that carbon from the pyruvate node could feed upwards in glycolysis/gluconeogenesis, and thus create a reduction imbalance, or it may be the cost of reduction power in the uptake that causes alanine to underperform.

Nitrogen sources giving growth but no product

Alanine, that was discussed above, is grouped with those nitrogen sources that supported growth, but did not yield a sufficient amount of product to be selected. Another amino acid in this group is Glutamine, which as sole nitrogen source performed less well than glutamate; likewise did asparagine perform less well than aspartate. Glutamine is directly interconvertible with glutamate requiring water, and releasing NH_3 . Considering the fact that the compounds were added as nitrogen sources, it could be speculated that glutamine and asparagine should be the more promising candidates, supplying 2 nitrogen molecules, as opposed to 1 from glutamate and aspartate. More so when comparing the amount of carbon each of the compounds contain relative to nitrogen. The reason why glutamine and asparagine were performing less well is yet to be determined; it could be related to uptake, or it could be the cost of reducing power, in the hydrolysis of the first ammonium. pH is unlikely as the cause, since the pK_a 's of all of the amino acids indicate no one of them should carry an unintended charge. A further observation that could be made was that glutamate performed better than aspartate, and glutamine performed better than asparagine – glutamate and glutamine are both C-5 compounds, whereas asparagine and aspartate are C-4. The difference in amount of growth and product that was achieved from these very similar compounds may be related to this extra carbon atom, giving the C-5 compounds a slight advantage over the C-4 compounds.

Further amino acids that were found in this group were: Cysteine, Glycine, Homoserine, Lysine and Serine, together with all of the peptides except for Ala-Gly. It is interesting that this group is so vast, since product formation should be expected to be closely correlated to growth, due to the product being expressed from a plasmid harboring a constitutive promoter. For cysteine, little has been reported on uptake and metabolization, but in *Bacillus spp.* it is metabolized via methionine (Burguière, et al., 2004).

D-Glutamate was also found in this group. D-glutamate is directly interconvertible with glutamate, using a racemase EC 5.1.1.3, for a non-energy requiring reaction, thus it shouldn't be expected that there would be a marked difference between L- and D-glutamate, though according to results from BRENDA deprotonation may be the rate limiting step, and thus slowing the further conversion of L-glutamate. There could also be a limitation in transport into the cells.

For quite a few of the amino acids, only little has been published, and it is not easy determining when a homologue can be expected. For Glycine nothing has been reported (Hodgson, 2000), though a decarboxylating dehydrogenase (from the glycine cleavage system) has been found on the genome of *S. coelicolor* (Bentley, et al., 2002), indicating that glycine can be used. Lysine is metabolised via cadavarine, (Hodgson, 2000), though it has been reported that *S. lividans* makes poor use of lysine as sole N-source (Madden, et al., 1996). Serine utilization is doubtful, the relatives *S. aureofaciens*, *S. rimosus* (Parada, 1981) and *S. clavuligerus* do not utilize serine as sole N-source (Hodgson, 2000).

A number of the sources that are located in the group with low-to-no production but a notable amount of growth are in the literature those that result in poor or no growth as single nitrogen sources in *S. lividans* or relative strains. This contradictory finding may be related to the nature of the Biolog plates, as opposed to the liquid cultures used in literature.

No growth – no production

Most of the amino acids are located in the “no-growth-no-production group”. When relating our results to what is known for streptomycetes, there is generally agreement. There are though some discrepancies; and some of our findings have not been published before. All of these sources produced too little growth and product to be worth investigating further in the scope of the paper. All of the sources marked with a star (*) are according to (Hodgson, 2000) not considered useful as sole nitrogen sources, and therefore show agreement to our results

The amino acids in this group are Methionine, Arginine, Phenylalanine *, Leucine*, Threonine*, Tryptophan*, Tyrosine*, Valine*, Isoleucine* and Proline (capable as metabolite in *S. coelicolor*) (Hodgson, 2000), as well as Ornithine and Citrulline. Voelker had a supporting observation, that branched chain amino acids, leucine, isoleucine and valine did not support growth when used as sole nitrogen sources (Voelker & Altaba, 2011).

The sources present in this group that are not in agreement with (Hodgson, 2000), are: Threonine, which it should be possible that *Streptomyces* could de-amine, and thus use. Proline can be metabolized by *S. coelicolor*, and would be expected to be used by *S. lividans*, though in this case we did not see growth, in either replicate. Tryptophan can be degraded, but, by most streptomycetes, not completely which results in accumulation of anthranilate and formyl-anthranilate. So while growth is possible, only a small amount can be expected when using either of the shikimate amino acids as sole nitrogen source. Little or nothing has been published on the degradation of methionine, and (Hodgson, 2000) mentions only one strain that can use methionine. For arginine, *S. coelicolor* can use this as single nitrogen source, and a GOGAT mutant created by random mutagenesis could not. *S. lividans* could grow on asparagine and alanine; therefore we can exclude the possibility that GOGAT is damaged, but there is still no apparent reason why arginine should not be used as sole nitrogen source.

Overall for the observations made in the Biolog plates, it is necessary to verify these findings in a more controlled medium. Especially the pH is difficult to control, and there is no verification what the pH actually is in the wells, so many of the discrepancies may be caused by pH related effects.

Scale up of experiments

In the scale-up to the 3 mL scale, some nitrogen sources did not result in the expected amount of growth. 2 of the initially selected nitrogen sources, pyroglutamic acid and histidine, did not support growth in the larger scale, according to literature, some *Streptomyces* should be capable of growing on histidine – however only a very slight growth was observed, which halted at very low biomass titers. The reason for the halted growth has not been found, despite Hodgsons reports that *S. coelicolor* is capable of growth on histidine (Hodgson, 2000; Kendrick & Wheelis, 1982).

Pyroglutamic acid, or 5-oxoproline is being degraded in glutathione metabolism, to glutamate, and can from here enter glutamate metabolism, or together with L-cysteine re-enter glutathione metabolism. The gene for 5-oxoprolinase (ATP-hydrolysing) [EC:3.5.2.9] is annotated in *Streptomyces avermilitis*, but apparently not in *S. coelicolor*, which in turn may mean that *S. lividans* also will not have it. There is a putative protein (ORF: SCO0627) in *S. coelicolor* that has some similarity to other known 5-oxoprolinases, and does contain signatures of an ATP binding cassette. Considering the lacking growth in a defined medium, may mean that the selection in the initial screen could have been an artifact. Further study, outside the scope of this paper, will be needed if it is to be verified whether either *S. coelicolor* or *lividans* possess the 5-oxoprolinase, and is indeed capable of metabolizing L-pyroglutamic acid. Allantoin as well as uric acid is degraded as part of the purine metabolism. In fact, uric acid is degraded to allantoin, and the final products in the pathway are ammonium and CO₂, via urea most likely, when looking at the KEGG annotations for *S. coelicolor*. However uric acid was not among the nitrogen sources that were selected, whether this is a transport matter or pH related effect is unknown.

Another compound that made the first round of selection, but was disregarded, was the Ala-Gly peptide. It is interesting to note that it solely of the peptides seems to perform well enough to be selected. It even outperforms its monomeric amino acids. Possibly this could be a benefit from a synergistic effect, as all peptides in this plate contain either Ala or Gly. All peptides scored med-high on growth, but not production. An option for speculation is that glycine by itself may enter gluconeogenesis, with glycine moving upwards in the glycolysis. Glycine, in conjunction with alanine, may on the other hand be forced into the TCA cycle, and therefore perform better. Whether it is the synergy of 2 favoured amino acids, a coincidence of a peptidase recognition sequence or an unbalanced amount of carbon that causes this peptide to score so well does remain unknown.

From previous experiments in large scale ammonium chloride and glutamate have been compared, and both perform well, and very similar to what was seen in these experiments. Ammonium has in some cases been reported to be a poor N-source resulting in a low growth rate, unlike our experience (Madden, et al., 1996). Using Nitrate though, results in high level of excretion of organic acids, pyruvate and 2-oxoglutarate (Madden, et al., 1996; Lee, et al., 1997), which can mostly be re-assimilated, but does it does affect growth. In general ammonium is considered the best N-source for *streptomyces*, followed by amino acids, and lastly nitrate.

Medium composition

The Biolog system is a commercial system, and the medium composition is proprietary, though similar to what is reported by (Bochner, et al., 2001). Further it seems as though the gelling agent may be metabolised, it is not defined what is being used, but in a patent they claim to use alginate, carrageenan and gellan gum (Bochner & Ralha, 1997), all sources that can be metabolised. To increase the comparability of our experiments, we have chosen to keep the amount of glucose and the total N-molarity of the nitrogen source fixed. The result is varying the total amount of carbon, but keeping the media as equal as possible, also across the scales, 3 mL – 1 L.

A balanced medium composition can be obtained where no unwanted limitation occurs; it is though difficult when comparing many chemically different N-sources. Some of the N-sources contain large amounts of carbon, in relation to the amount of nitrogen. The amino acids, glutamate and aspartate have a 5:1 and 4:1 C:N ratio respectively. That means that comparing these with ammonium and nitrate, which contain one nitrogen and no carbon per molecule, does give the amino acids an advantage, since they have more available carbon. Having chosen a smaller subset of nitrogen sources, the next likely study would be to compare those, under conditions where the C:N ratio is kept constant. This may be done by either adding a further N-source, so as to balance the amount of carbon supplied by the N-source, or by accepting that the final amount of glucose will change between the experiments.

Final remarks

The general application of choosing a number of nitrogen sources to screen for increased production, with RFP as a model protein yielded not only a number of candidates showing good productivity, but by virtue of using the Biolog system yielded an astounding amount of physiological information. Coupling these findings to larger volumes, and more defined parameters confirmed the mutual growth/productivity pattern, as in the small volumes. The screening principle demonstrated fitness, and has the potential to be taken further to a more automated selection. A possible further step could be to find good combinations of carbon/nitrogen sources, or even complete media compositions based on such a screening.

References

- Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., et al. (2002, May). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417(6885), 141-147.
- Binnie, C., Cossar, J. D., & Stewart, D. I. (1997, Aug). Heterologous biopharmaceutical protein expression in *Streptomyces*. *Trends Biotechnol*, 15(8), 315-320.
- Bochner, B. R., Gadzinski, P., & Panomitros, E. (2001, Jul). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res*, 11(7), 1246-1255.
- Bochner, B., & Ralha, M. (1997, May). Multi-test format with gel-forming matrix for characterization of microorganisms, U.S. Pat. 5,627,045
- Borodina, I., Schöller, C., Eliasson, A., & Nielsen, J. (2005, May). Metabolic network analysis of *Streptomyces tenebrarius*, a *Streptomyces* species with an active entner-doudoroff pathway. *Appl Environ Microbiol*, 71(5), 2294-2302.

- Brawner, M. E. (1994, Oct). Advances in heterologous gene expression by *Streptomyces*. *Curr Opin Biotechnol*, 5(5), 475-481.
- Burguière, P., Auger, S., Hullo, M.-F., Danchin, A., & Martin-Verstraete, I. (2004, Aug). Three different systems participate in L-cystine uptake in *Bacillus subtilis*. *J Bacteriol*, 186(15), 4875-4884.
- D'Huys, P.-J., Lule, I., Hove, S. V., Vercammen, D., Wouters, C., Bernaerts, K., et al. (2011, Apr). Amino acid uptake profiling of wild type and recombinant *Streptomyces lividans* TK24 batch fermentations. *J Biotechnol*, 152(4), 132-143.
- D'Huys, P.-J., Lule, I., Vercammen, D., Anné, J., Van Impe, J. F., & Bernaerts, K. (2012, Sep). Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. *J Biotechnol*, 161(1), 1-13.
- Hodgson, D. A. (2000). Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol*, 42, 47-238.
- Jain, R., Agrawal, S. C., & Jain, P. C. (2009). Proteolytic Actinomycetes from Indian Habitats. *Journal of Culture Collections*, 6(No 1), 28-37.
- Kendrick, K. E., & Wheelis, M. L. (1982, Sep). Histidine dissimilation in *Streptomyces coelicolor*. *J Gen Microbiol*, 128(9), 2029-2040.
- Keränen, S., & Penttilä, M. (1995, Oct). Production of recombinant proteins in the filamentous fungus *Trichoderma reesei*. *Curr Opin Biotechnol*, 6(5), 534-537.
- Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). Practical streptomyces genetics. John Innes Foundation.
- Lammertyn, E., Mellaert, L. V., Schacht, S., Dillen, C., Sablon, E., Broekhoven, A. V., et al. (1997, May). Evaluation of a novel subtilisin inhibitor gene and mutant derivatives for the expression and secretion of mouse tumor necrosis factor alpha by *Streptomyces lividans*. *Appl Environ Microbiol*, 63(5), 1808-1813.
- Lee, M. S., Kojima, I., & Demain, A. L. (1997, Aug). Effect of nitrogen source on biosynthesis of rapamycin by *Streptomyces hygroscopicus*. *J Ind Microbiol Biotechnol*, 19(2), 83-86.
- Macauley-Patrick, S., Fazenda, M. L., McNeil, B., & Harvey, L. M. (2005, Mar). Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, 22(4), 249-270.
- Madden, T., Ward, J. M., & Ison, A. P. (1996, Nov). Organic acid excretion by *Streptomyces lividans* TK24 during growth on defined carbon and nitrogen sources. *Microbiology*, 142 (Pt 11), 3181-3185.
- Mahmoud, K. (2007). Recombinant Protein Production: Strategic Technology and a Vital Research Tool. *Research Journal of Cell and Molecular Biology*, 1, 9-22.
- Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., & Porro, D. (2012). Recombinant protein production in yeasts. *Methods Mol Biol*, 824, 329-358.

- Morello, E., Bermúdez-Humarán, L. G., Llull, D., Solé, V., Miraglio, N., Langella, P., et al. (2008). *Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J Mol Microbiol Biotechnol*, 14(1-3), 48-58.
- Parada, J. L. (1981, Feb). Growth Inhibition of *Streptomyces* Species by L-Serine and Its Effect on Tetracycline Biosynthesis. *Appl Environ Microbiol*, 41(2), 366-370.
- Pozidis, C., Lammertyn, E., Politou, A. S., Anné, J., Tsiftoglou, A. S., Sianidis, G., et al. (2001, Mar). Protein secretion biotechnology using *Streptomyces lividans*: large-scale production of functional trimeric tumor necrosis factor alpha. *Biotechnol Bioeng*, 72(6), 611-619.
- Punt, P. J., Schuren, F. H., Lehmbeck, J., Christensen, T., Hjort, C., & van den Hondel, C. A. (2008, Dec). Characterization of the *Aspergillus niger* prtT, a unique regulator of extracellular protease encoding genes. *Fungal Genet Biol*, 45(12), 1591-1599.
- Punt, P. J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J., & van den Hondel, C. (2002, May). Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol*, 20(5), 200-206.
- Sianidis, G., Pozidis, C., Becker, F., Vrancken, K., Sjoeholm, C., Karamanou, S., et al. (2006, Feb). Functional large-scale production of a novel *Jonesia* sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*. *J Biotechnol*, 121(4), 498-507.
- Sohoni, S. V., Bapat, P. M., & Lantz, A. E. (2012). Robust, small-scale cultivation platform for *Streptomyces coelicolor*. *Microb Cell Fact*, 11, 9.
- Strickler, J. E., Berka, T. R., Gorniak, J., Fornwald, J., Keys, R., Rowland, J. J., et al. (1992, Feb). Two novel *Streptomyces* protein protease inhibitors. Purification, activity, cloning, and expression. *J Biol Chem*, 267(5), 3236-3241.
- van den Hombergh, J. P., van de Vondervoort, P. J., Fraissinet-Tachet, L., & Visser, J. (1997, Jul). *Aspergillus* as a host for heterologous protein production: the problem of proteases. *Trends Biotechnol*, 15(7), 256-263.
- van Ooyen, A. J., Dekker, P., Huang, M., Olsthoorn, M. M., Jacobs, D. I., Colussi, P. A., et al. (2006, May). Heterologous protein production in the yeast *Kluyveromyces lactis*. *FEMS Yeast Res*, 6(3), 381-392.
- van Wezel, G. P., Krabben, P., Traag, B. A., Keijser, B. J., Kerste, R., Vijgenboom, E., et al. (2006, Aug). Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering. *Appl Environ Microbiol*, 72(8), 5283-5288.
- Vitikainen, M., Hyyryläinen, H.-L., Kivimäki, A., Kontinen, V. P., & Sarvas, M. (2005). Secretion of heterologous proteins in *Bacillus subtilis* can be improved by engineering cell components affecting post-translocational protein folding and degradation. *J Appl Microbiol*, 99(2), 363-375.
- Voelker, F., & Altaba, S. (2001, Sep). Nitrogen source governs the patterns of growth and pristnamycin production in '*Streptomyces pristinaespiralis*'. *Microbiology*, 147(Pt 9), 2447-2459.

- Vrancken, K., & Anné, J. (2009, Mar). Secretory production of recombinant proteins by *Streptomyces*. *Future Microbiol*, 4(2), 181-188.
- Westers, L., Westers, H., & Quax, W. J. (2004, Nov). *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim Biophys Acta*, 1694(1-3), 299-310.
- Zhang, H., Wang, Y., Wu, J., Skalina, K., & Pfeifer, B. A. (2010, Nov). Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem Biol*, 17(11), 1232-1240.

Figure 1 Scatterplot-Plot of Area under curve (AUC) values of growth and fluorescence, averaged between two plates, colored according to the groupings obtained by hierarchical clustering

Figure 2 A venn diagram displaying how many N-sources were selected by growth, fluorescence or both

Figure 3 Growth rate versus specific production rate for the 3-mL scale and reactor experiments. Histidine and pyroglutamic acid did not support growth at and are left out. The filled symbols are from the reactor experiments, the hollow are from the microtiter. Diamond is Allantoin, square is ammonium, triangle is aspartate, circle is nitrate, the bar is glutamate, and only represented in the microtiter scale. All growth rates, and production rates are relative to ammonium, therefore the hollow square is on top of the filled square in (1,1)

Figure 4a,b,c,d - Representations of fermentation time profiles for (a) ammonium, (b) aspartate, (c) allantoin and (d) nitrate. Biomass measured in (g/L) is marked by green triangles, glucose measured in (g/L) is marked purple crosses, mRFP measured in (g/L) is marked with blue stars (for nitrate, the concentration of mRFP has been multiplied by 10), CO₂ measured in volume % is marked by the blue line.

Table 1 Amount of nitrogen source, to supply 100mM N

N-source	g/L
Ammonia	5.35
Sodium nitrate	8.49
L-AsparticAcid	13.31
L-GlutamicAcid	18.71
L-Histidine	6.99
D-GlutamicAcid	14.71
L-PyroglutamicAcid	12.91
Allantoin	3.95

Figure 1 - Scatterplot

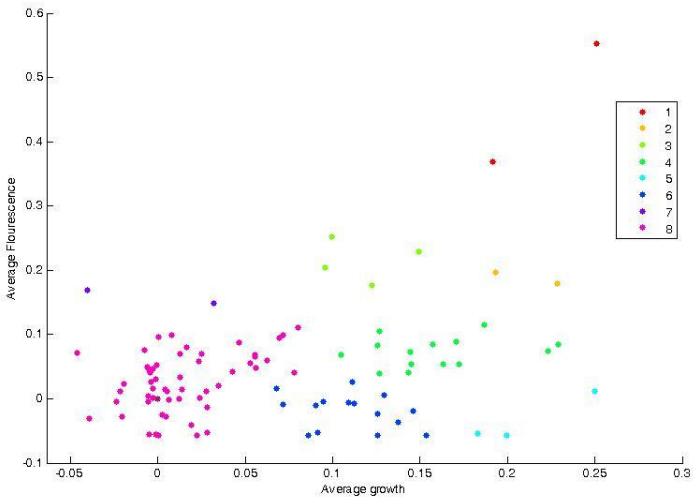


Figure 2 – Venn diagram

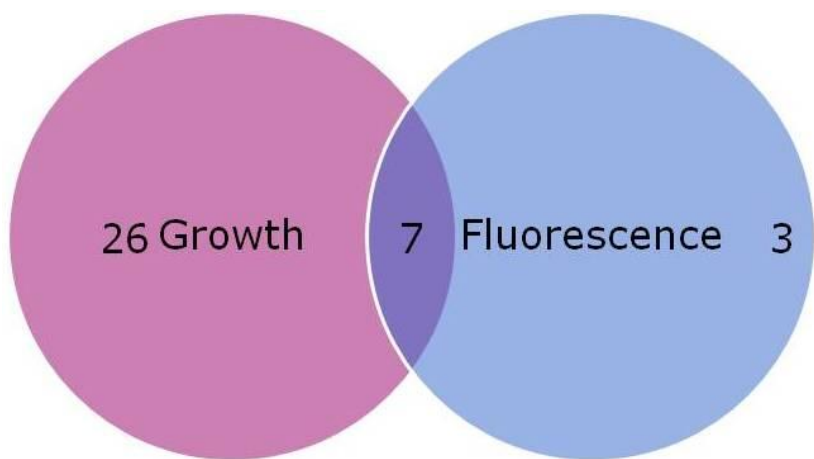


Figure 3

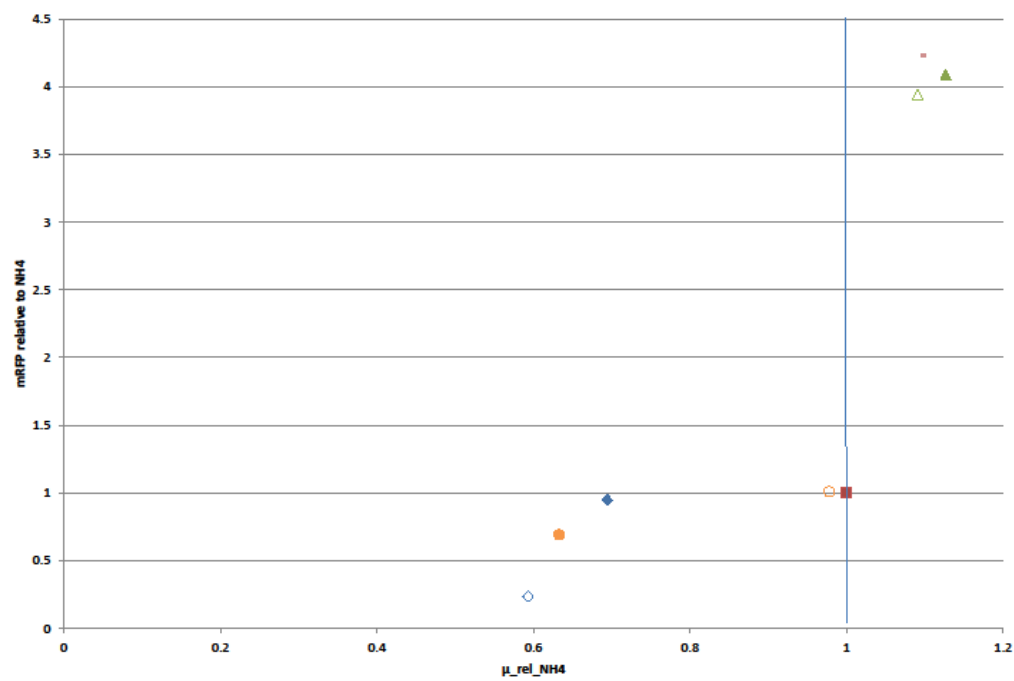


Figure 4a

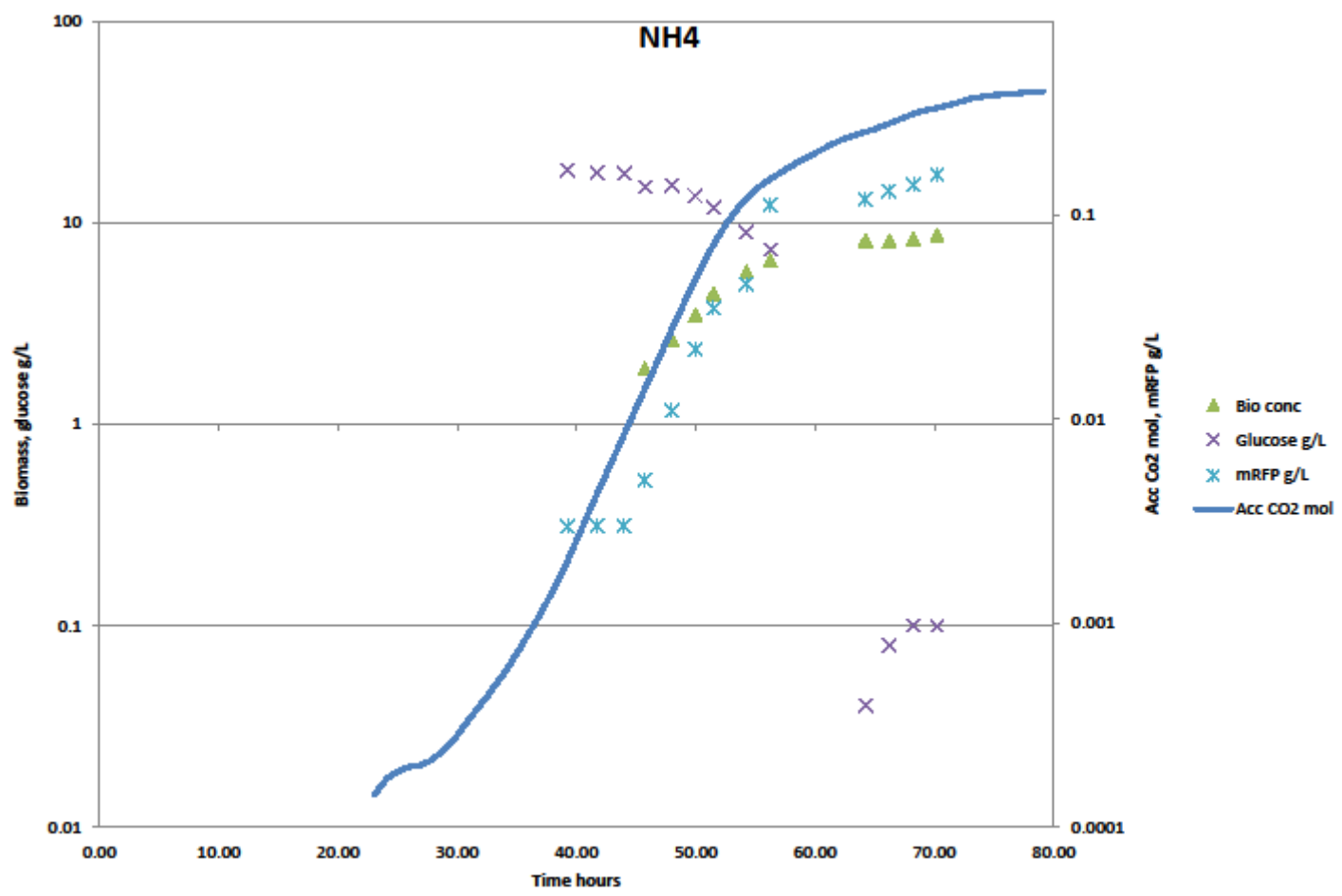


Figure 4b

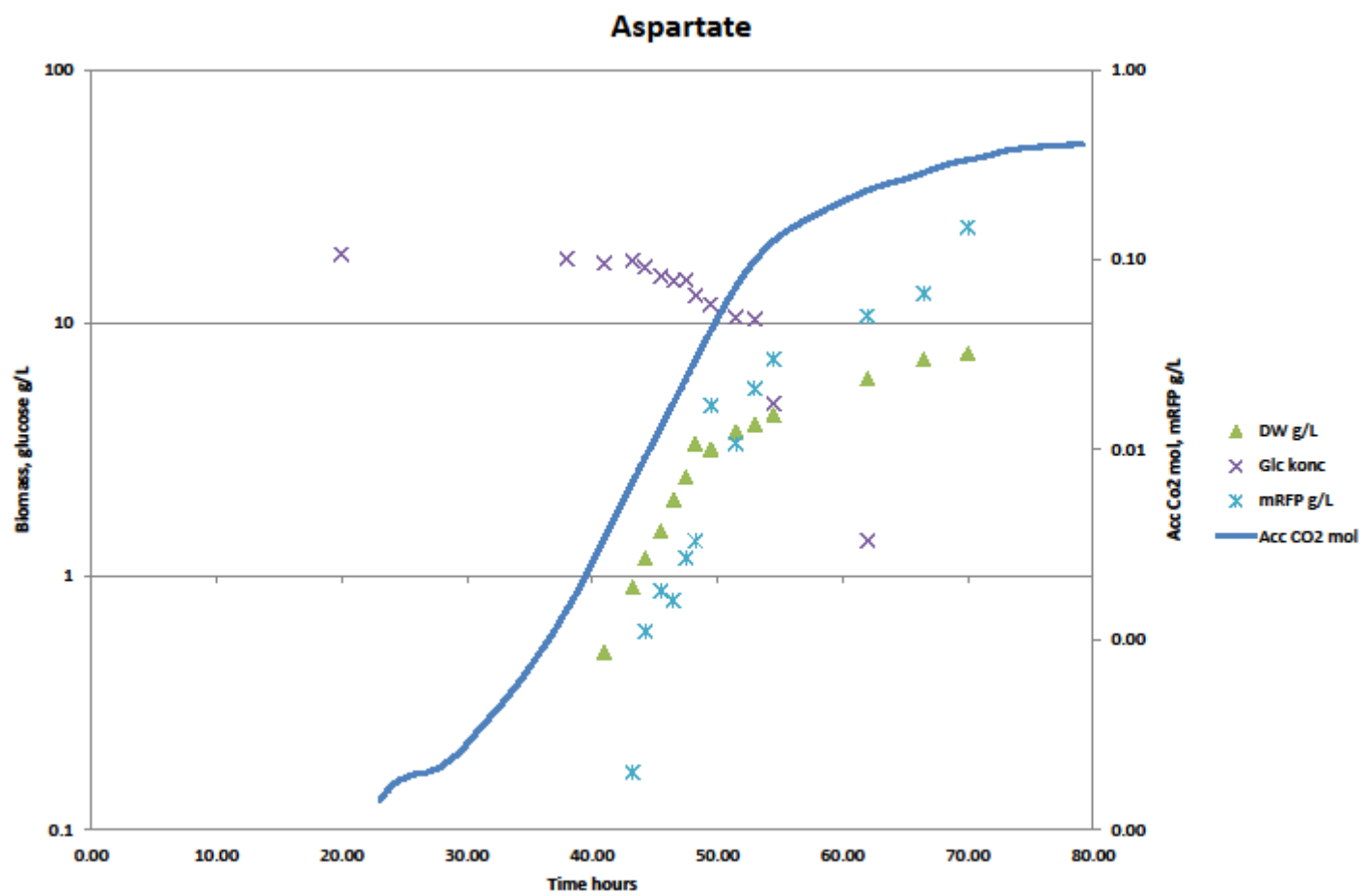


Figure 4c

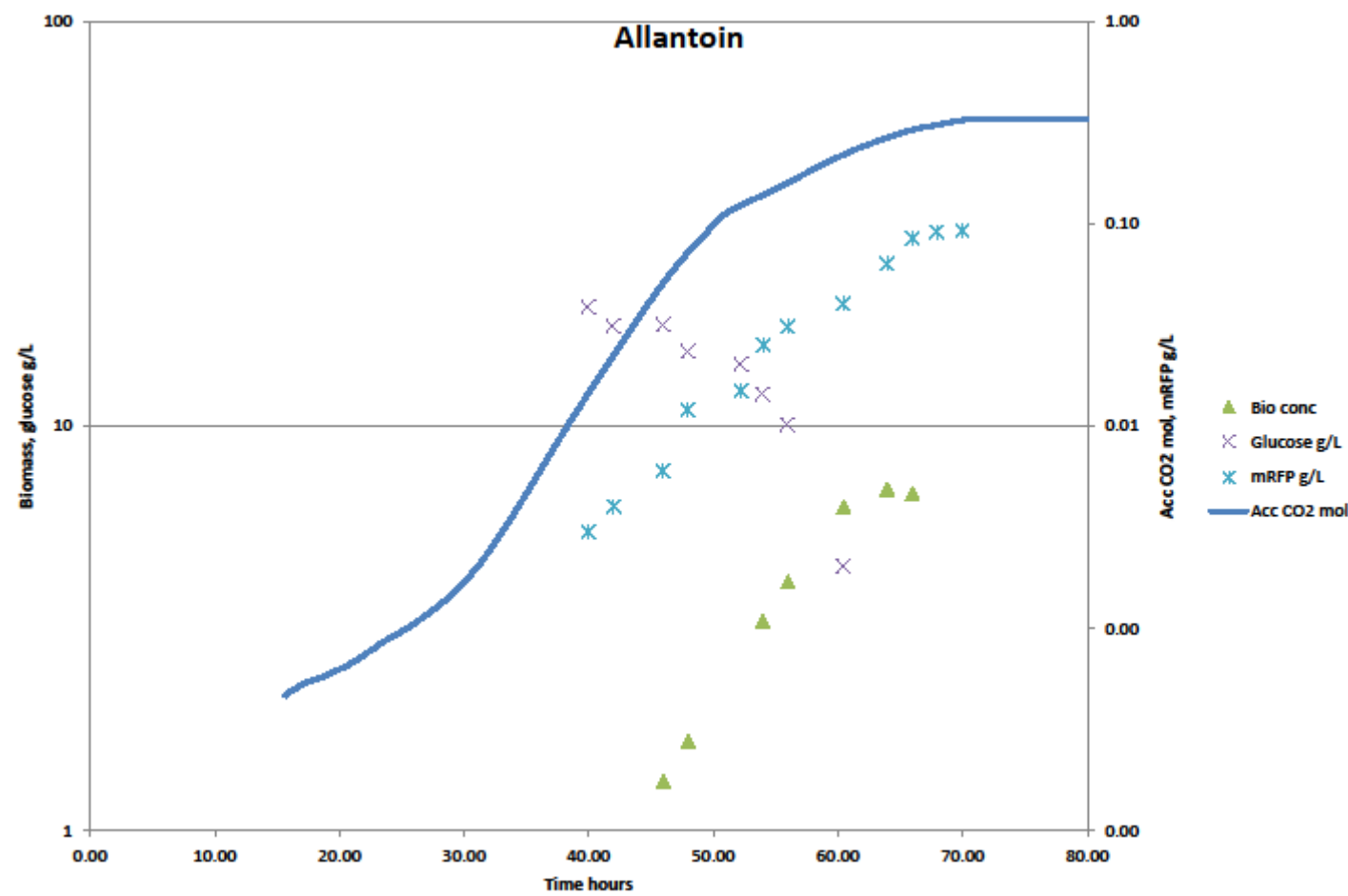
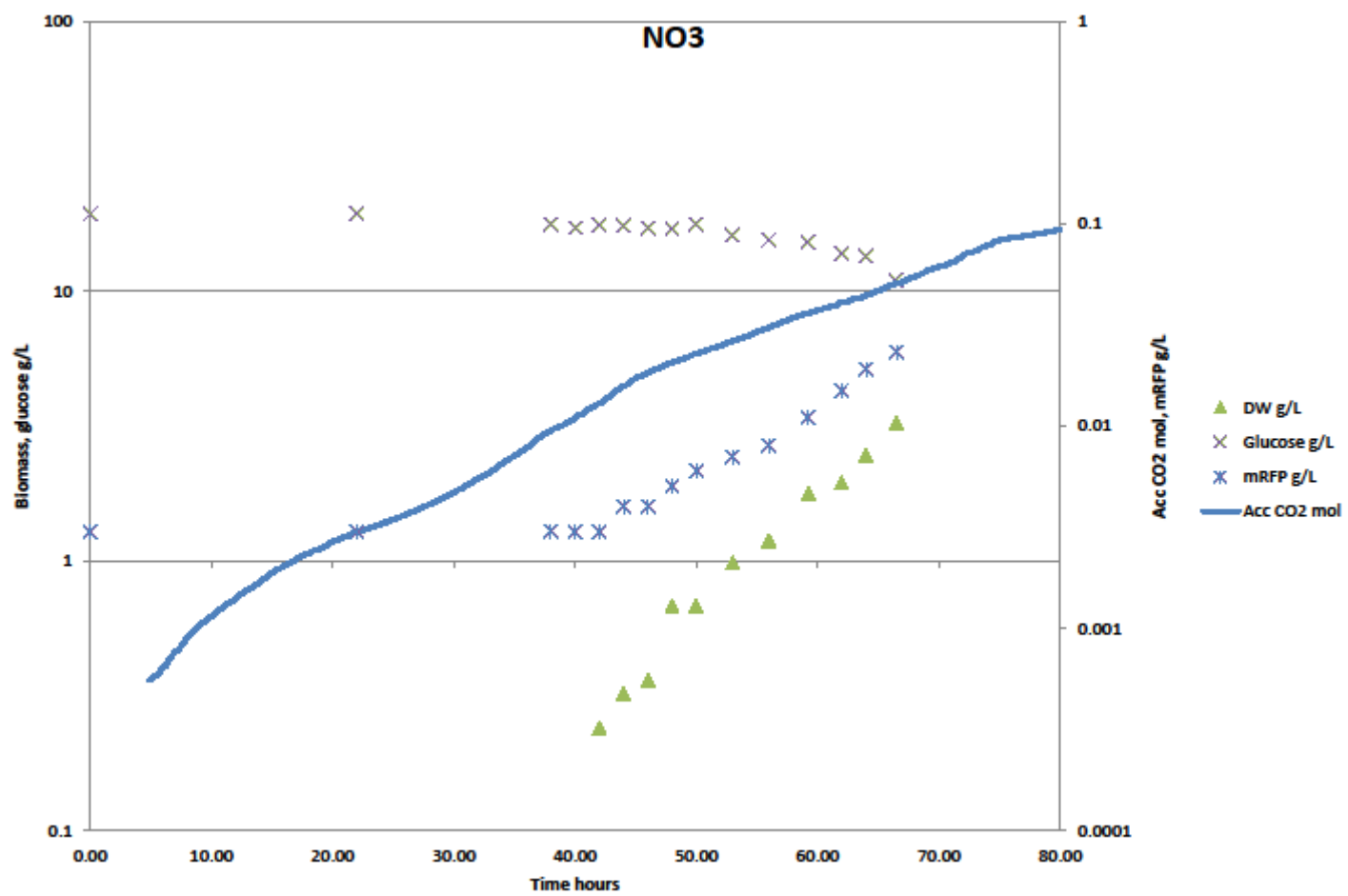


Figure 4d



Chapter 5. Chemostats

Fermentation

As a term fermentation historically refers to the anaerobic conversion of substrates by microorganisms. With time the meaning, when considering the process, has changed to encompass any kind of controlled cultivation of microorganisms; aerobic and anaerobic alike. Many products are produced through fermentation, and more recently, many industrial products, including solvents that can be polymerised to plastics have become part of the repertoire. The most basic type of fermentation is the batch process, which in the very simplest of cases literally is putting substrates in a pot, waiting a bit (to gather wild microorganisms), then sealing the pot; a method that has been used in beer brewing since the middle ages. Industrially less reliance is placed on luck, and a thorough sterilisation scheme is used, as well as monoseptic inoculation. Batch fermentations are characterised by having a high concentration of substrates, and through most of the process a low concentration of biomass, making the process susceptible to effects like catabolite repression (both carbon and nitrogen), as well as resulting in a low final titer of the product. Batch processes have been used in chapters 3 and 4 to test the initial performance of *S. lividans* as heterologous host, and impact of nitrogen source on heterologous protein production. Due to catabolite repression the fed batch was invented, as it was discovered that by slow addition of sugar to a fermentation of *Penicillium chrysogenum*, penicillin could be produced without being limited by carbon catabolite repression. After the initial batch, the growth rate in a Fed-batch is controlled by the feeding rate, depending on the scheme chosen. The schemes available are exponential, linear, and constant. The exponential profile is chosen to extend the exponential growth phase, though this method is severely limited by mass transfer issues, both of medium addition, but also oxygen supply etc. The linear profile, which is technically much simpler than the exponential, limits the growth rate to a linear profile, but allows continuous development of biomass under non catabolite repressing conditions. Lastly, the constant profile is akin to the continuous culture, except for the dynamic volume. Industrially all types of processes are used, including a repeated-fed-batch, which is run as a fed-batch, mostly emptied, and fed again in a semi-continuous manner. Lastly, the continuous process is a setup that requires the continuous addition and removal of medium. Many specific subtypes of continuous cultures have been developed each with their specific purposes; some seek to push dilution rates as high or higher than μ_{\max} (Hoskisson & Hobbs, 2005), among these are the retentiostat, where biomass is separated from the waste stream and fed back into the reactor (Bull, 2010), allowing $D > \mu_{\max}$. In other cases very low dilution rates are of interest, in the literature doubling times of more than 700 h have been reported for continuous cultures of *Saccharomyces cerevisiae* (Boender, et al., 2009). Other examples are the auxostat types of continuous culture, where an external parameter is used to control the feed addition rate, e.g. pH or optical density (OD) (Bull, 2010). There are also the serially linked processes to impose different limitations in each step, e.g. serially lower and lower growth rate. In the following I will limit myself to the chemostat, which is characterised by a fixed volume, and constant concentrations in time, at steady state.

Chemostat

The continuous culture, in our case more specifically the chemostat is advantageous when examining the conditions that lead to optimal productivity, and they have been used as such since the sixties. One of the preconditions, and major benefits, of the chemostat is that the growth rate of the organism is limited by a

single substrate, all other substrates are in excess, and this allows the precise control of the growth rate of the organism (Hoskisson & Hobbs, 2005). Chemostats can be used for studies of genetic stability, strain evolution, physiology, and regulatory mechanisms. Further with the advent of systems biology tools and 'omics' technologies, the chemostat can be used to reliably, repeatedly, and reproducibly generate biomass for advanced 'omics' analyses under different conditions (Bull, 2010). One thing that is important to keep in mind is that while the chemostat was originally designed to study cultures at steady state as defined by Monod, it also offered the unprecedented opportunity to study accelerated evolution, as the very high selective pressure in the chemostat causes forced adaptation (Ferenci, 2006). Limiting a single nutrient does cause a starvation response, and it can be argued whether a chemostat is ever in an actual steady state, since the residual nutrient concentration will continue to decrease over time (Ferenci, 2006). The usefulness of the chemostat in generating controlled repeatable experiments is so pronounced that accepting a 'quasi' steady state with the proper precautions, considering the before mentioned potential side effects when studying e.g. the metabolome, transcriptome or proteome, still warrants its use (Hoskisson & Hobbs, 2006).

The general chemostat setup physically differs from a batch reactor in that feed is constantly added, and spent medium, containing cells for most processes, is constantly removed. The way in which a chemostat differs from a continuous culture, or can be considered as a specific sub-type of continuous cultures, is that the chemostat has no time dependent developments, at steady state. In reality mutations are prone to occur, and could give a subpopulation an advantage that will allow it to take over the culture (Ferenci, 2006) (Hoskisson & Hobbs, 2005) (Hoskisson & Hobbs, 2006) (Bull, 2010); the selective advantages in a chemostat would likely relate to substrate specificity.

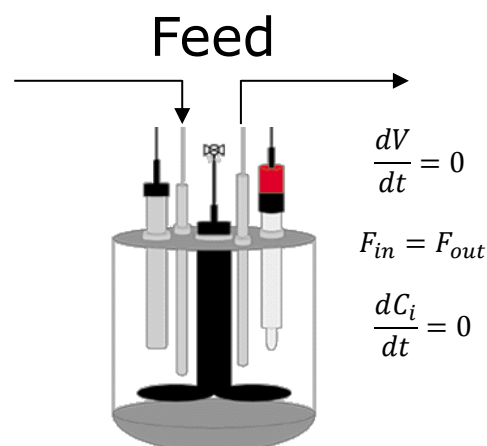


Figure 14 A depiction of a general chemostat setup. What is special about the chemostat, compared to other types of continuous culture, is that the volume (V) is constant, and at steady state all concentrations are constant. For V to be constant, $Flow_{in}=Flow_{out}$. The concentration of the limiting substrate in the $Flow_{out}=0$.

For the setup of the experiment it is required to choose one limiting substrate, typical choices are carbon, nitrogen or phosphorous, as exemplified by glucose, ammonium or phosphate respectively; though any medium component, as long as it is growth rate limiting, can be used. The experiment is started as a normal

batch, to build up biomass concentration. When a sufficient biomass concentration has been reached, and the limiting nutrient is almost expended, the pumps can be started for exchanging the medium. In batch the cells will grow at their, for the given conditions, maximal growth rate (μ_{max}). In a chemostat, the constant exchange of medium is intended to cause the convergence of a steady state. In essence the growth rate of the organism is being controlled by controlling the flow rate of fresh medium addition. This medium has the single nutrient limitation, and therefore excess of all other nutrient components. When a steady state is achieved, all parameters are constant in time. The relation between growth and substrate can be given by the Monod equation:

$$\mu = \mu_{max} * \frac{C_s}{C_s + K_s}$$

C_s is the substrate concentration, μ the growth rate, μ_{max} the maximal growth rate, determined in the batch, and K_s is the substrate specificity, or half velocity constant, found as the substrate concentration when $\mu = \frac{1}{2} * \mu_{max}$. The change in biomass over time in any continuous process can be defined as:

$$\frac{dx}{dt} = \mu x - D(x_{out} - x_{in})$$

D is the dilution rate x_{in} is biomass going in with the feed, for a chemostat this value is 0, x_{out} is the biomass sampled from the out flow (equal to the concentration in the vessel). Since the change in biomass with time at steady state is 0, the result is:

$$0 = \mu x - D(x_{out} - 0) \Leftrightarrow \mu = D$$

$$D = \frac{F}{V}$$

F is flow rate in, and V is the liquid volume in the vessel. Since the definition of the chemostat requires the volume to be fixed, then it is clear that the growth rate μ , is equal to D , the dilution rate, and that it is controlled as a function of the flow rate.

Previous studies with filamentous bacteria

Studies in chemostats can besides physiological characterisation and determining the optimal growth productivity relation (Fazeli, et al., 1995) also be used to estimate physiological factors, such as genetic stability and investigate regulation of various limitations. The environment in a chemostat is as mentioned before highly selective, and depending on the type of limitation, mutations are prone to arise, or plasmids can be lost. One study of genetic stability of *S. lividans* in a chemostat was performed by Roth in 1985 (Roth, et al., 1985), where the cultivation lasted for more than 100 generations. This study showed that apparently carbon and phosphate limitation impose the strongest selective pressure on *S. lividans*, as independent of growth rate, these conditions caused plasmid loss faster than observed for nitrogen limitation. The growth rate (D in the range $0.1-0.2h^{-1}$) seemed to only have a limited impact on plasmid loss, but temperature had a more pronounced impact; temperatures of $28^{\circ}C$ and $36^{\circ}C$, with the higher temperature causing the loss.

Most available physiological studies of production in *Streptomyces* are related to antibiotics. Although some features can be learnt from these, they generally require a nutrient limitation imposed on the organism, which induces secondary metabolism, in order to promote antibiotic production. Phosphate has been reported as the nutrient of choice to limit if the desire is to produce antibiotics (Bushell, et al., 2006). This seems reasonable since phosphate is required to be incorporated in the cell wall, and that the environment in the soil is reported to be phosphate and nitrogen limited (Hodgson, 2000). In Table 2 the studies are referenced where *S. lividans* has been cultivated in a continuous culture. The list may not be perfectly exhaustive, though it represents the studies that could be found, as well as their purpose.

Table 2 A list of papers where chemostats or continuous cultures are used with *Streptomyces lividans*.

Name	Limitation	Purpose
(Roth, et al., 1985)	C/N/P	Genetic stability
(Noack, et al., 1988)	N	Basic physiology – Mutant selection / Heterologous protein
(Undisz & Noack, 1993)	C	Genetic stability
(Lee & Lee, 1994)	C	Genetic stability / heterologous protein production
(Parro, et al., 1998)	Mg / PO ₄	Phosphate limitation on heterologous protein (CCR?)
(Isiegas, et al., 1999)	P	Glucose repression? Heterologous protein
(Wang, et al., 1998)	Complex	Heterologous protein
(Sarrà, et al., 1998)	P	Antibiotic
(Rossa, et al., 2002)	P	Flux – antibiotics

Having determined that the literature on *S. lividans* as a heterologous host is broad, but contains few physiological studies, the next step is to shed some light on this area of *S. lividans* physiology. What is sought in this physiological study, is investigating the connection between growth rate and productivity, as well as study how *S. lividans* performs as a producer of heterologous proteins.

Physiological studies

In batch it is possible to investigate the highest possible growth rate (μ_{max}), and by extension the highest Y_{xp} (yield of product on biomass). Many products are not directly growth related, or in the case of secondary metabolism, have an inverse relationship between productivity and growth. In the chemostat it can be investigated how the connection is between growth and productivity (Fazeli, et al., 1995), as well as comparing strains under exactly the same conditions, even growth rate, since $\mu=D$, which is the fundamental basis of the chemostat. There is a limit to the range of growth rates that can readily be used in a chemostat, the D_{crit} is the upper limit of flow rate that can be used, since anything beyond this causes washout, or requires change of the experimental setup to a retentiostat.

Carbon is a common limitation to impose for heterologous protein production; the benefits include limiting the risk of activating secondary metabolism and limiting the effect of catabolite repression. Carbon is straightforward to limit, since it is among the most abundant elements in the biomass. If the medium is of the defined minimal type, then there should only be one carbon source. Whereas it is possible to limit

some of the less common nutrients and minerals, it makes higher requirements for buying of chemicals; e.g. magnesium limited chemostats (Parro, et al., 1998) and (Utkilen, 1982). Some of these nutrients can be growth limiting, albeit they are required at low concentrations. The experiment by (Utkilen, 1982) found that in batch, μ_{\max} was reached with addition of 5 μ M magnesium, thus at this concentration the limitation was overcome.

Strains and media

The strain used was *Streptomyces lividans* TK24. All proteins, mRFP and CelA are expressed from the plasmid pIJ486 (Kieser, et al., 2000), which is a multicopy plasmid bearing a Thiostrepton resistance marker, as well as the sequence for the protein behind the strong *vsj* promoter and signal peptide of the *Streptomyces venezuelae* subtilisin inhibitor (Sianidis, et al., 2006). The CelA-19 strain (will be referred to as Cel-A) was kindly supplied by Olafur Friðjónsson (Anné, et al., 2012). The mRFP strain was kindly provided by the lab of Prof. Jozef Anné, Catholic University Leuven. The medium used for germination was 2YT containing in g/L Yeast extract 16, Tryptone 10 and NaCl 5 (Kieser 2000). Solid sporulation medium is MS agar containing in g/L Mannitol 20, Soy flour 20, agar 20. Spore stocks were prepared according to (Kieser, et al., 2000), in a final concentration of glycerol of 20%vol.

For growth in liquid culture a defined/minimal medium was used containing in g/L NaH₂PO₄·H₂O 2.76, NH₄Cl 5.35, KCl 0.75, Na₂SO₄ 0.28, Citric acid·H₂O 0.42, MgCl₂·6H₂O 0.25, CaCl₂·2H₂O 0.18, Trace metal sol 5 mL/L, Antifoam (sigma 204) 0.2 mL/L Adjust volume 0.9 L. After autoclavation add Vitamin 1 mL/L and separately autoclaved Glucose·H₂O 20 – giving 100 mM glucose. The medium composition is adapted from Evans (Borodina, et al., 2005)

Trace metal solution has the following compositions in g/L: FeCl₃·6H₂O 5.406, CuCl₂·2H₂O 1.705, ZnCl₂ 6.815, MnCl₂·4H₂O 1.98, Na₂MoO₄·2H₂O 0.00484, CoCl₂·6H₂O 4.76, and H₃BO₃ 0.618.

Vitamin solution g/L: Biotin 0.05, Ca-pantpthenate 1, nicotinic acid 1, myo-inositol 25, thiamine-HCl 1, pyridoxine-HCl 1, para-aminobenzoic acid 0.2.

Thiostreptone was added to the medium to a final concentration of (5mg/L)

Fermentation experiments

Two fermentation setups have been used:

- 1) Sartorius 1L bioreactors controlled by BioPAT towers. Temperature was 30°C, stirring 600 rpm with two 6-bladed rushton impellers; pH was kept at 6.8+/-0.1 by addition of either 2M NaOH or 2M H₂SO₄. Aeration was adjusted to 1VVM. Off-gas was recorded using a Prima PRO Process Mass Spectrometer (Thermo Fisher Scientific), capable of measuring ethanol, methanol, O₂, CO₂, Nitrogen, Argon etc, and autocalibrating regularly.
 - a. Chemostats where run using gravimetric feed control connected to an analog controlled Watson-Marlow pump. Level control was weight based, using the digitally controlled “anti-foam” pump on the BioPAT tower
- 2) Applikon 1L reactors, with two 6-bladed Rushton impellers. Stirring was set to 600 rpm, pH was adjusted to 6.8+/-0.1 by addition of either 2M NaOH or 2M HCl. The pH controller was capable only

of controlling two channels, thus one per fermentor. Aeration was set to 1 VVM, using a mass-flow controller. Temperature was adjusted to 28°C using an open water bath with manual temperature control. DO was measured using a Mettler-Toledo unit. Off gas was measure using a Brüel & Kjær acoustic gas analyser, capable of measuring CO₂ and O₂. The setup is presented in Figure 15.

- a. Chemostats where run using constant feed on a manually calibrated Watson-Marlow pump. Level control was weight based, using a digitally controlled Watson-Marlow pump. The feed bottle was standing on a scale, where weight was manually noted at regular intervals, to verify the feed rate.

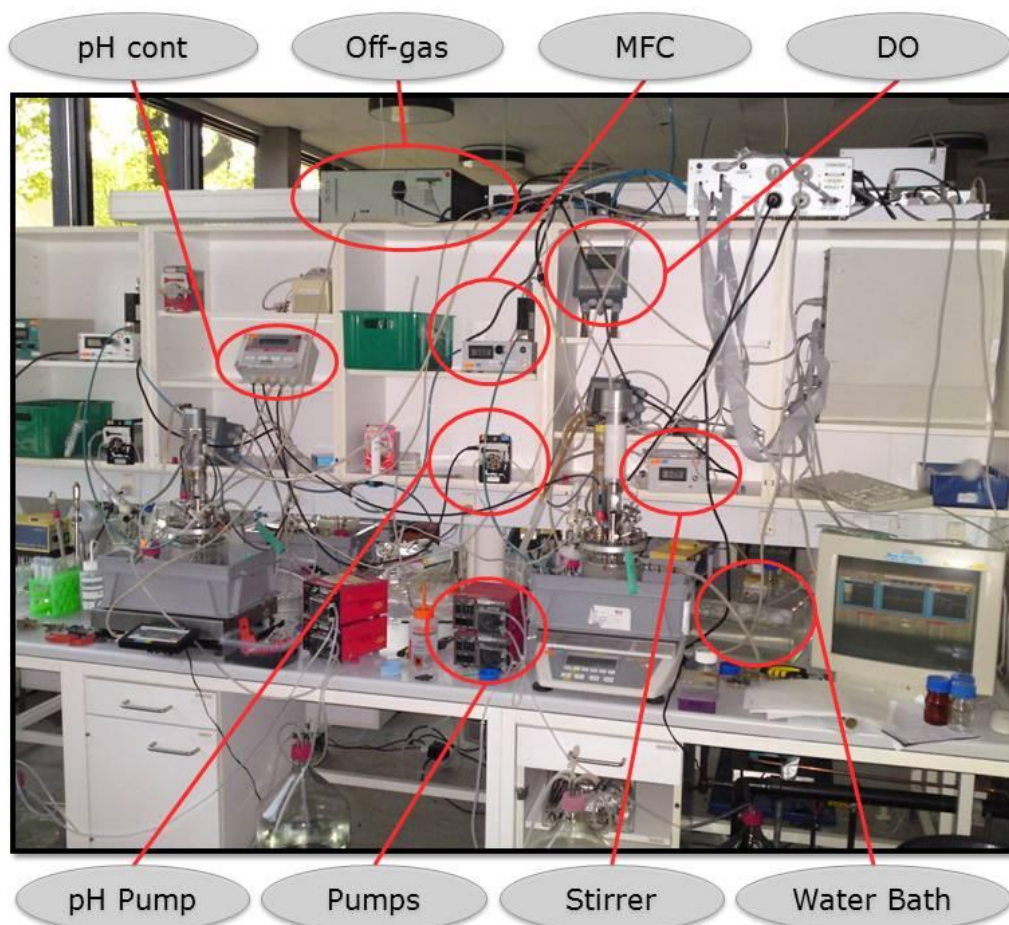


Figure 15 A picture of the Applikon setup of chemostats. The picture is added for historical purposes.

Samples were taken for biomass dry weight, extra-cellular metabolites and enzymatic activity analyses.

Dry weight sampling was performed by filtrating 5 mL culture through a 0.45µm PES membrane (Sartorius Stedim, Germany) using a vacuum pump. The supernatant was collected and stored at -18°C for later analysis of glucose and mRFP concentrations. The dry weight filters were dried in a microwave oven before and after filtration for 20 min. at 150W – allowed to cool in a dessicator, and weight was noted before and after respectively.

Glucose is measured using a glucose dehydrogenase kit (ABX pentra Glucose HK CP, Hobira Montpellier, France).

mRFP concentration is obtained by measuring fluorescent activity in a Biotek plate reader, using 50µL culture, and an mCherry protein from Biovision as standard. For fluorescent measurement black-sided 96 well plates from Greiner are used. The filters used in the Biotek reader are excitation 530/25nm and emission 595/35nm

Results

The strains investigated are the WT and the mRFP producing strain. First step is to determine whether it is possible to establish a steady state, because *Streptomyces* as mentioned grows filamentous, which can cause large heterogeneity and oscillations of the culture. Morphologically it can be quite challenging, since part of the population of cells can grow as pellets (as the wild type does), which can cause mass transfer limitations. Even if establishing something that resembles a stable condition, the absolute stability from a steady state can be eluding, e.g. due to plasmid loss, or variance in mass transfer (see (Sarrà, et al., 1998)).

To enable chemostat cultivations a single limiting nutrient has to be chosen to fit the desired experiment. Few have reported chemostats at all in *S. lividans*, most are not C-limited, but rather phosphate or nitrogen limited, and interested in studying other aspects of physiology, among others antibiotic production.

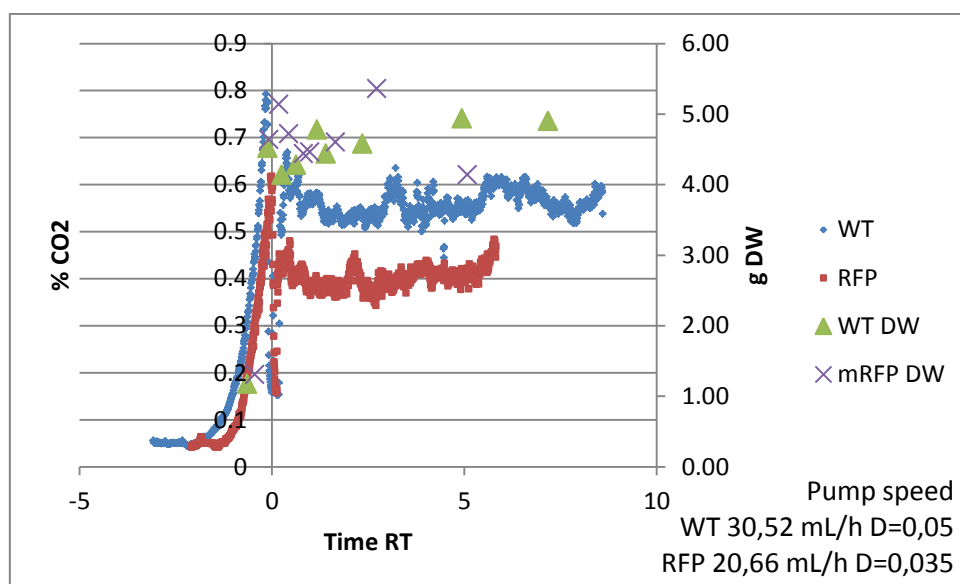


Figure 16 Establishing a steady state

As can be seen from Figure 16 a steady production of CO₂ can be initiated from almost the time when the feed is added. The continuous culture depicted in Figure 16 is run in the setup showed in the picture in Figure 15. The setup was somewhat more rudimentary than what was later installed, though for all intents and purposes it worked admirably. As can be seen from Figure 16 the WT strain adjusted to a steady state, after 3-5 residence times. It started by growing in large dense pellets, but as the growth was limited by the feed, the morphology opened up, and the strain started growing freely dispersed. The mRFP strain on the other hand, was growing freely dispersed throughout the cultivation. It changed colour, and became blue,

due to actinorhodin production. After the change it did not maintain biomass concentration or obtain a new steady state, within the timeframe of the experiments. As can also be seen from Figure 16, the dilution rates are 0.05h^{-1} and 0.035h^{-1} for the WT and mRFP strain respectively. The aim was to obtain a D in the scale of $0.04\text{--}0.05\text{h}^{-1}$, as this was half the μ_{max} of the very initial experiments; choosing half of μ_{max} allows for room for error in estimation of μ_{max} , and in operating the equipment. The duration of each experiment in Figure 16 is equally long, but because of the lower dilution rate of the mRFP strain, the amount of residence times were less. It also seems that for the mRFP strain it was the very low dilution rate that resulted in starving of the cells and caused the onset of the secondary metabolism; there is no other indication from the data that any other perturbation took place. This sets the lower limit for running a chemostat somewhere between $D=0.035\text{h}^{-1}$ and $D=0.05\text{h}^{-1}$, assuming the primary purpose is to avoid secondary metabolism, and secondly to investigate heterologous protein production. Having a protein that is expressed from a plasmid using a constitutive promoter does not exclude production of protein concurrently with the secondary metabolism, though a major shift of the metabolism would suggest that less carbon would be expected to be directed towards protein production, as well as when antibiotics are produced there are more stringent requirements for waste disposal.

With an unsuccessful run of the mRFP strain, a higher dilution rate was attempted. This next experiment was performed in the newer reactor setup with the Sartorius reactors. This means that the level and the feed control both are gravimetric, in contrast to the older setup where the feed was constant, based on precalibrating the pump together with the tubing and later verifying the rate by noting the weight change of the feed bottle manually. This makes the newer system less labour intensive

The benefits of using a classical pre-calibrated system is that it is quite stable, though it will tend to drift slightly over time, as the tubing gets worn in the peristaltic pump. The main con is that it is very labour intensive setting up. The gravimetric feed pump is not susceptible to air bubbles and the PID will set the pump speed based on the feed weight, which prevents drifts in dilution rate over time. Unfortunately my experience is that it takes longer to reach a stable pump speed, since the PID has a tendency to start at max pumping speed, and then gradually reduce the speed, which causes pulsing addition in the early part of the experiment. The gravimetric system, while precise, is more susceptible to fluctuations, and risky if using a shared lab. Very importantly, when using a gravimetric system all wires and tubing should be fixed as efficiently as possible, since even small changes in how the tubing drags, or lies on the table, can affect the final volume of the weight based system, as well as the dilution rate.

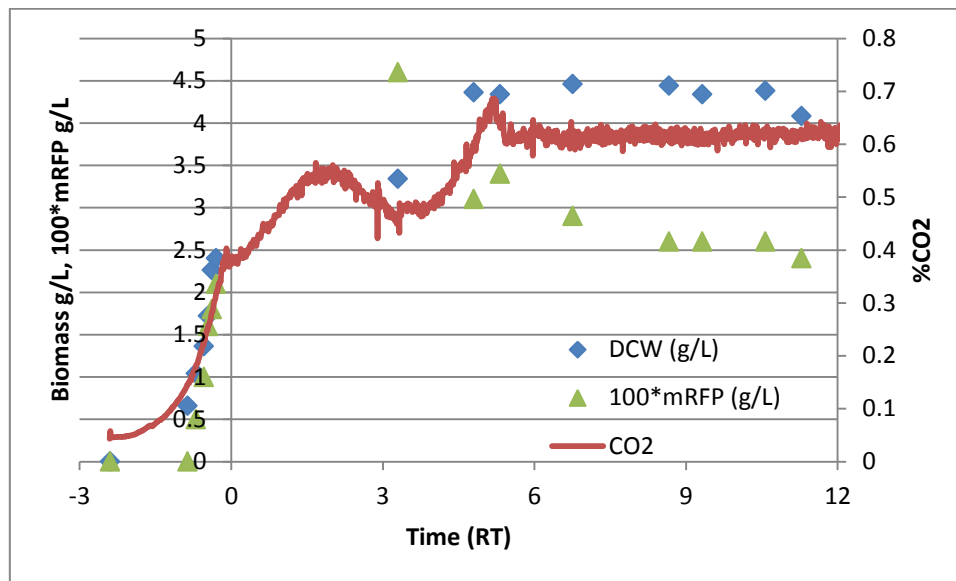


Figure 17 Testing if it is possible to establish a steady state in a chemostat at $D=0.08h^{-1}$

In Figure 17 an experiment is presented, with only the mRFP strain, this time with a dilution rate $d=0.08h^{-1}$. The experiment has been performed in the newer Sartorius bioreactors, coupled to the BioPAT towers. It can be seen that the adaptation phase is somewhat longer, but a steady state is found, and maintained. Surprisingly almost 6 residence times were needed to obtain a steady state, once this was attained though, the system was very stable. It seems the long adaptation was caused by beginning of the feeding phase too soon, which resulted in the culture not having enough biomass to quickly adapt, and thereby extending the time in rapid growth. For comparison, when running yeast in chemostat, it is conventional practice to start the feed after CO₂ drops, just after the exponential phase when glucose is expended; if however this method was followed for the *Streptomyces* strains we used, it was observed that secondary metabolism was initiated, and a stable steady state was not obtained. Therefore we found that it was necessary to observe the residual glucose concentration by regularly sampling using a glucose stick, usually used to test for glucose in urine of diabetes patients. The stick has a maximum limit of 1000 mg/dL = 10 g/L – where it is dark green, and then with different graduations towards “normal” which is less than 500 mg/L. Once less than 1 g/L is left the pumps can safely be started.

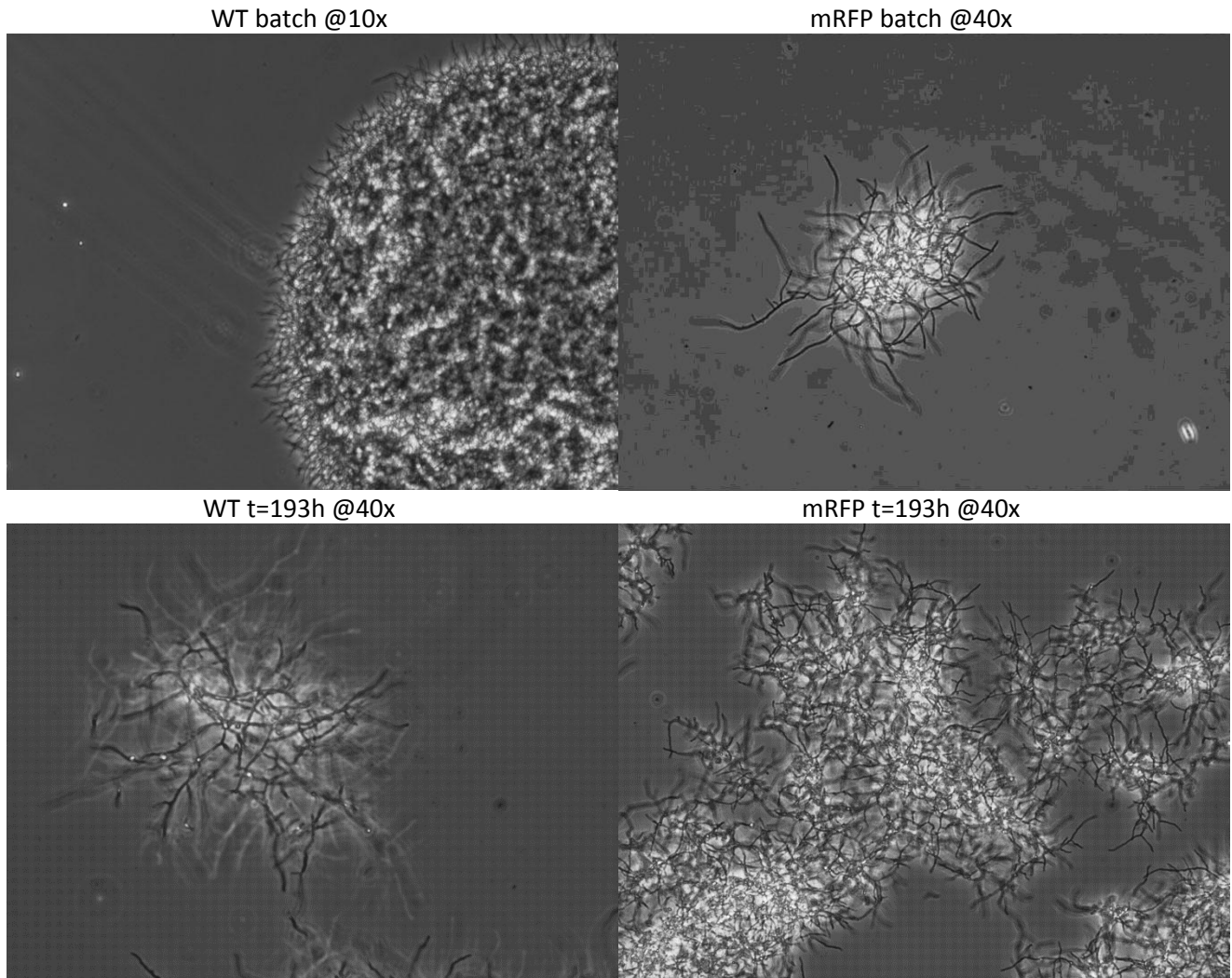


Figure 18 Microscopy pictures of chemostat experiments, WT on the left, and the mRFP producing strain on the right. The two upper pictures are from the batch phase, and the two lower pictures are from the steady state.

In Figure 18 microscopy pictures from the chemostat experiment are presented, and it was observed that the WT strain, which grew pelleted in batch, started to loosen up in its morphology and eventually became freely dispersed. The pellets of the WT strain can become quite large, and as can be seen the microscopy pictures below, the pellets are very dense. The RFP strain on the other hand grows freely dispersed all the way.

Investigating the productivity in chemostats

Having determined that it was possible to establish a steady state, the relationship between growth and productivity was to be explored. For doing this, three dilution rates were chosen: $D = 0.04$, 0.08 and 0.12h^{-1} .

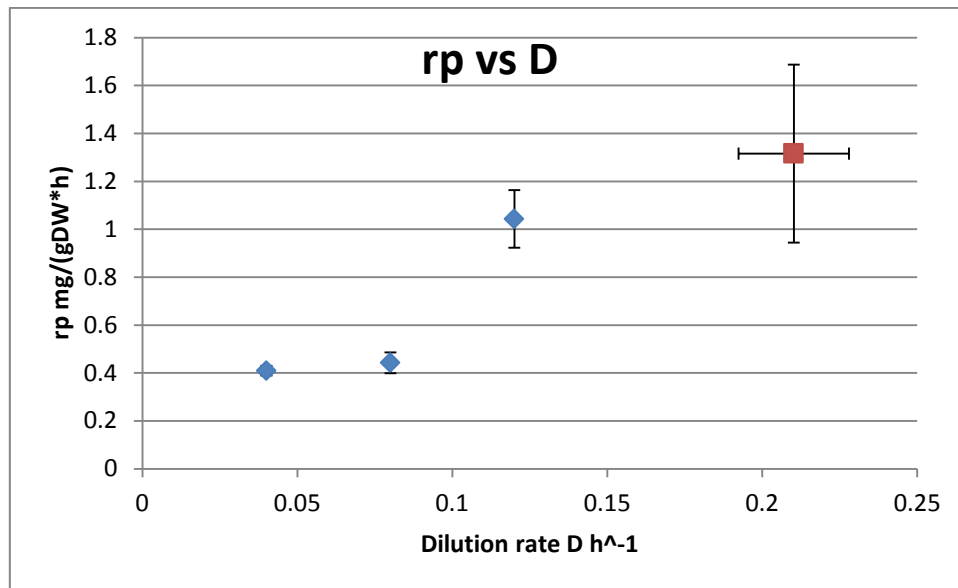


Figure 19 rp vs D – specific rate of productivity plotted against D - the square is the batch results plotted to extend the range to include μ_{\max}

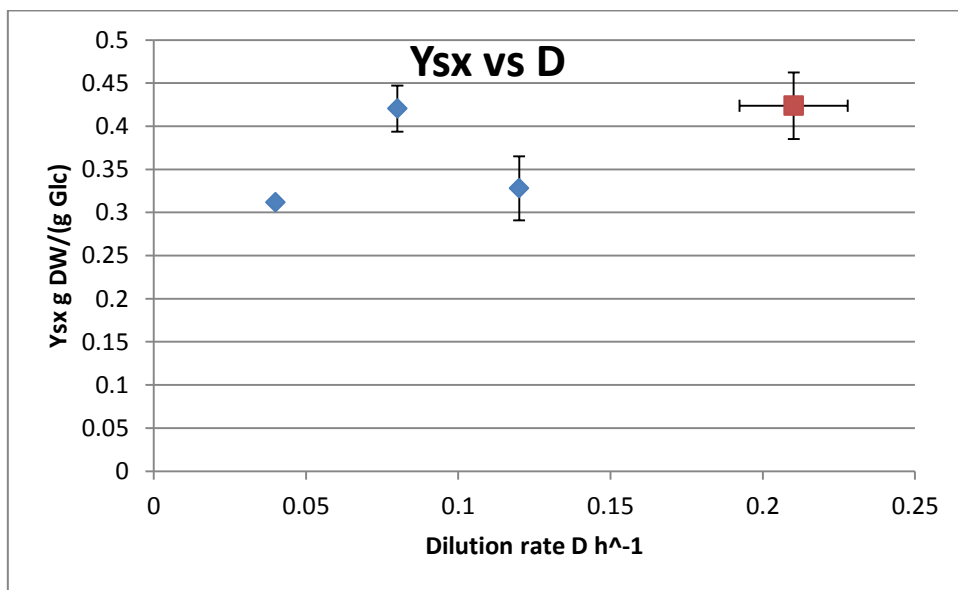


Figure 20 Ysx vs D – yield of biomass on substrate plotted against D, the square is the batch results plotted to extend the range to include μ_{\max}

As can be seen in Figure 19 there is a relationship between growth and productivity that resembles linearity. This is expected, as the product is transcribed from a constitutive promoter on a plasmid, and as long as cellular machinery is not saturated, the correlation should be linear. It is not obvious why $D=0.04 \text{ h}^{-1}$ and $D=0.08 \text{ h}^{-1}$ separate so little in terms of productivity. When including the productivity data from the batch (red square), the results suggest that there is room for further optimization of production, since there seem to be no excessive burden of production on the strain, and it seems $D=0.08 \text{ h}^{-1}$ should have had a higher rp. Using the data presented in Figure 20, the yield of biomass on substrate in relation to D, it can indeed be suggested, that it is the $D=0.08 \text{ h}^{-1}$ that underperforms, in relation to productivity. The Ysx of $D=0.08 \text{ h}^{-1}$ is at batch level, indicating a partial loss of plasmid.

Discussion

In order to relate our results to those of the previously performed studies, the elements of these will be discussed shortly below. From Table 2 it was seen that most of the previous studies performed on *S. lividans*, relating to continuous culture, are intent on studying genetic stability or regulation. While these are useful studies, the amount of studies that are dedicated to investigate the heterologous protein production and characterisation of the producing strain are few. Two are concerned with glucose repression on the specific product, and one uses a complex nutritional source.

One of the studies with carbon limitation was performed by (Roth, et al., 1985), and here the chemostat experiments were started by running a prolonged cultivation at $D=0.1\text{h}^{-1}$, from which a mutant strain was selected that had developed a free mycelial structure, and retained this even after making a stock, and starting new experiments; this strain could be cultivated at $D=0.2\text{h}^{-1}$. It is interesting, though very difficult to determine, what mutation the strain had picked up, and as such it makes it difficult to directly compare the findings of Roth, though this group does present the best experiment. Further, Roth uses the *S. lividans* 66 strain, which is the parental strain of the TK24, with an integrated fertility plasmid (Lewis, et al., 2010).

(Rossa, et al., 2002) used a P-limited medium to investigate flux distribution in an *S. lividans* strain producing the antibiotics actinorhodin and undecylprodigiosin (UDP). One finding that was made was that there is an inverse relationship between antibiotic productivity, and growth rate.

The two papers of Mellados group (Isiegas, et al., 1999) and (Parro, et al., 1998) claim to use chemostats to investigate heterologous expression. Though it is undefined why magnesium limitation is chosen as the reference condition. The chemostats are run, with either glucose or mannitol as carbon source, first on magnesium limitation, and after 5 residence times shifted to phosphate limitation. Unfortunately no full fermentation profiles are displayed for verification of the steady states. It seems as though, but it is not really defined as a goal, that magnesium limitation is used with intent to retain a high residual glucose/mannitol concentration in the vessel, so that phosphate limitation can be compared to another condition with high residual carbon. When shifting to the phosphate limited medium not only phosphate limitation is tested, but also carbon catabolite repression because in both cases a high residual carbon concentration is maintained, while controlling the growth rate with another nutrient source. It seems unfortunate choosing Mg^{2+} as the limiting nutrient as this could have a host of secondary/indirect effects, not unlike those that was attempted to be avoided by not adding the selective antibiotic; despite that the argumentation of avoiding secondary effects was exactly the reason for not adding the selective marker. Further, not running with a selective marker is a risky proposition, as Roth showed that exactly phosphate limitation seems to induce plasmid loss. Lastly Mg^{2+} is used in a large number of enzymes, among others in transport and glycolytic enzymes. And the paper by Isiegas is based on the same promoter, and same assumptions, though a different protein product.

(Wang, et al., 1998) is studying expression of a heterologous protein in *S. lividans*, though the experiment is performed in a continuous culture using a rich medium. This may be very interesting in terms of producing the enzyme, though in academia, running a continuous culture using a complex medium, as trypticase soy broth is, makes for an outcome, with very hard to interpret results. It could also potentially result in different limitations depending on batch variation of the complex medium. (D'Huys, et al., 2011) has described the amino acid uptake, in a parallel experiment, where the complex part was supplied by

casamino acids, Though D'Huys studied only batch experiments, and later made steady state inferences from late exponential phase flux, but still from batch data.

While the setup chosen by (Sarrà, et al., 1998) seems to be well thought through, the extremely low dilution rates that were chosen also become the limitation of the system. With a dilution rate of $D=0.014\text{h}^{-1}$ and $D=0.007\text{h}^{-1}$ result in residence times of 71 and 143 hours respectively. For the duration depicted where the experiment was set to $D=0.014\text{h}^{-1}$ it can be observed that antibiotic production ceased. After four residence times, where biomass seemed steady, the dilution rate was shifted to $D=0.007\text{h}^{-1}$ to attempt to restart antibiotic production. The duration depicted that was chosen for running $D=0.007\text{h}^{-1}$ is 150h, thus only one residence time, and therefore not enough to establish and verify a steady state. Normal convention requires at least three residence times for establishing a steady state, and the steady state samples from three consecutive residence times, to verify the steady state. Further, (Sarrà, et al., 1998) mentions a Glutamic acid limitation, while feeding both another carbon and nitrogen source. While glutamic acid is a favored amino acid, it can supply mainly carbon and nitrogen, and if other sources are available, it can't be an ultimately growth limiting substrate.

Having considered the work of these other groups, how has development been in terms of *Streptomyces lividans* as a heterologous protein producing strain. We have established, that it is possible to perform chemostat experiments with *S. lividans*, and that a range of dilution rates are available. There seem to be a level, slightly faster than $D=0.12\text{h}^{-1}$, where at least, a two-step chemostat will have to be performed in order to obtain reliable steady states at the higher dilution rates, as in the experiments of Roth. This is despite using the mRFP strain with a $\mu_{\text{max}} \approx 0.22\text{h}^{-1}$ which should easily allow dilution rates of 0.16h^{-1} , if not higher, though this was not successful. From a genetic stability perspective, there are some problems with plasmid stability. A number of chemostats have been performed, where either the addition of thiostrepton has been forgotten, or the stock of thiostrepton has been faulty, and here it is very clear that the plasmid can be mostly lost within a short timeframe. I say mostly because the mRFP production does not cease entirely, and after all it is a multicopy plasmid with 40-100 copies per cell (Ward, et al., 1986); a coordinated loss of that many plasmid copies seems unlikely.

Works Cited

- Anné, J. et al., 2012. Recombinant protein production and streptomycetes.. *J Biotechnol*, Apr, 158(4), pp. 159-167.
- Boender, L. G. M. et al., 2009. Quantitative physiology of *Saccharomyces cerevisiae* at near-zero specific growth rates.. *Appl Environ Microbiol*, Sep, 75(17), pp. 5607-5614.
- Borodina, I., Krabben, P. & Nielsen, J., 2005. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism.. *Genome Res*, Jun, 15(6), pp. 820-829.
- Bull, A. T., 2010. The renaissance of continuous culture in the post-genomics age.. *J Ind Microbiol Biotechnol*, Oct, 37(10), pp. 993-1021.
- Bushell, M. E., Kirk, S., Zhao, H.-J. & Avignone-Rossa, C. A., 2006. Manipulation of the physiology of clavulanic acid biosynthesis with the aid of metabolic flux analysis. *Enzyme and Microbial Technology*, 39(1), pp. 149-157.
- D'Huys, P.-J. et al., 2011. Amino acid uptake profiling of wild type and recombinant *Streptomyces lividans* TK24 batch fermentations.. *J Biotechnol*, Apr, 152(4), pp. 132-143.
- Fazeli, M. R., Cove, J. H. & Baumberg, S., 1995. Physiological factors affecting streptomycin production by *Streptomyces griseus* ATCC 12475 in batch and continuous culture.. *FEMS Microbiol Lett*, Feb, 126(1), pp. 55-61.
- Ferenci, T., 2006. A cultural divide on the use of chemostats.. *Microbiology*, May, 152(Pt 5), pp. 1247-1248.
- Hodgson, D. A., 2000. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria.. *Adv Microb Physiol*, Årgang 42, pp. 47-238.
- Hoskisson, P. A. & Hobbs, G., 2005. {Continuous culture - making a comeback?}. *Microbiology*, 151(10), pp. 3153-3159.
- Hoskisson, P. A. & Hobbs, G., 2006. One size doesn't fit all.. *Microbiology*, Jul, 152(Pt 7), p. 1891.
- Isiegas, C., Parro, V. & Mellado, R. P., 1999. *Streptomyces lividans* as a host for the production and secretion of *Escherichia coli* TEM beta-lactamase.. *Lett Appl Microbiol*, Apr, 28(4), pp. 321-326.
- Kieser, T. et al., 2000. *Practical streptomyces genetics*. s.l.:John Innes Foundation.
- Lee, J. H. & Lee, K. J., 1994. Effect of growth rate and cultivation environments on cloned gene stability and the cloned gene product formation in *Streptomyces lividans*.. *J Biotechnol*, Mar, 33(2), pp. 195-204.
- Lewis, R. et al., 2010. Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. *BMC Genomics*, 11(1), p. 682.
- Noack, D. et al., 1988. Expression and secretion of Interferon- β by *Streptomyces lividans*: use of staphylokinase signals and amplification of a neo gene. *Gene*, 68(1), pp. 53-62.

- Parro, V., Mellado, R. P. & Harwood, C. R., 1998. Effects of phosphate limitation on agarase production by *Streptomyces lividans* TK21. *FEMS Microbiology Letters*, 158(1), pp. 107-113.
- Rossa, C. A. et al., 2002. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*. *Metab Eng*, Apr, 4(2), pp. 138-150.
- Roth, M., Noack, D. & Geuther, R., 1985. Maintenance of the recombinant plasmid *pIJ2* in chemostat cultures of *Streptomyces lividans* 66 (*pIJ2*). *Journal of Basic Microbiology*, 25(4), pp. 265-271.
- Sarrà, M., Pérez-Pons, J. A., Gòdia, F. & Casas Alvero, C., 1998. Importance of growth form on production of hybrid antibiotic by *Streptomyces lividans* TK21 by fed-batch and continuous fermentation. *Appl Biochem Biotechnol*, 75(2-3), pp. 235-248.
- Sianidis, G. et al., 2006. Functional large-scale production of a novel *Jonesia* sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*. *J Biotechnol*, Feb, 121(4), pp. 498-507.
- Undisz, K. & Noack, D., 1993. Autonomous selection of differentiation mutants of *Streptomyces lividans* TK24 in continuous culture. *Biotechnology Letters*, Årgang 15, pp. 13-18.
- Utkilen, H. C., 1982. Magnesium-limited Growth of the Cyanobacterium *Anacystis nidulans*. *Journal of General Microbiology*, 128(8), pp. 1849-1862.
- Wang, F., Whitaker, R. D. & Batt, C. A., 1998. Production of glucose isomerase in a recombinant strain of *Streptomyces lividans*. *Applied Microbiology and Biotechnology*, Årgang 50, pp. 65-70.
- Ward, J. M. et al., 1986. Construction and characterisation of a series of multi-copy promoter-probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase gene from Tn5 as indicator. *Mol Gen Genet*, Jun, 203(3), pp. 468-478.

Chapter 6. Beyond the black box

The black box

The experiments described in the previous chapters that were run in either batch or chemostats are in general black-box experiments; they are simple representations of fermentation processes. The substrates that are going into the cell are known, external parameters are controlled, and products coming out of the cells can be measured. That is to say, we know what goes in and what comes out, but we do not know exactly how the substrates are distributed into metabolites in the cell. Black box modelling is the basic approach when considering modelling, since no *a priori* knowledge is required and it is based on all the measurable parameters around the cell, and even the cell concentration; but we still work with the assumption that we don't need to know precisely what is going on inside the cell, in order to get the product out. Depending on which school of modelling one belongs to, some would claim black-box modelling requires the biomass to be considered as a single component. Using black box modelling, and suitable experimental conditions, much can be learned about the biology of the cell, including revealing parts of the network structure. The restraints made on black box models are measurable extracellular parameters. Flux analysis is one way of trying to estimate what is happening intracellularly, though it requires either that the network structure is simple or additional information to constrain the network in order to give a reasonable estimation of the carbon distribution. One of the most often used methods to simplify the network structure is by lumping reactions, in e.g. linear pathways, or all reactions going to biomass. These simplifications work because using a defined medium, all components are known, so consumption of medium components will result in either accumulation of biomass or a metabolite; this can be simplified to what goes in must come out. Due to the amount of reactions present in the cellular network compared to the parameters that can be readily measured, the system is vastly underdetermined (Kauffman, et al., 2003). Biotechnology in general has developed around this paradigm, we can attempt to describe what goes on by considering the cells as black boxes; we can isolate and purify intracellular enzymes for characterisation *in vitro*, and thereby develop a picture of the possible reaction network, though knowledge about what precisely happens in the cell is still limited by our ability of prediction (Palsson, 2002). We have learned to experiment and draw conclusions based on what can be measured. It is only relatively recently in the use of microorganisms that the genomic sequences have become known in their entirety, and to a great extent has been used to support and extend the reaction networks; further, the analytical techniques have undergone vast improvements and are becoming sufficiently advanced in both analysing small amounts of analytes, as well as separate the many compounds in the cell (analysing many compounds at a time).

Beyond the Black box

For initial experiments black-box modelling are fine. Many kinetic models function using extracellular parameters, such as glucose uptake rate, yield of product on substrate, yield of biomass on substrate, growth rate etc. Models like this can be used, after initially estimating the parameters, to predict the outcome of fermentations. The models can become quite robust, and the even the somewhat variable nature of the *Streptomyces* is not beyond the ability of the models (Nandy, et al., 2013). The models can even be expanded to include multiple substrates, each with their own specificity, and from experimental

data the specificities can be determined (Gernaey, et al., 2010). The parameters of the black box model describes the cellular performance very well, but eventually, if the strain is intended for use as a heterologous protein production platform, a deeper knowledge of the intracellular environment is needed. The kinetic models, while very useful, have the inherent weakness that they describe a known strain, under known conditions, and thus cannot be used to predict the effect of a different genotype (loss or gain). If improvements are to be made on the cellular physiology, then as the first step the reaction network needs to be clarified, preferably in its entirety. Intracellular networks can help to elucidate possible optimization regimes that may contain unexpected, unobvious, or even counter intuitive solutions (King, 1997). As an example of this, knockout of the TTA recognising tRNA (*bldA*) in *S. coelicolor*, has been shown to have effects even on genes in the primary metabolism, despite the expectation that this particular codon was confined to secondary metabolism (Hesketh, et al., 2007).

Overall there are two methods for reconstructing genome scale networks, either a bottom up-, or a top down approach (Palsson, 2002) (De Keersmaecker, et al., 2006). The top down approach is trying to define the network, without *a priori* knowledge, from the data alone; a large amount of data, such as transcription data from a large number of states, is necessary for this to succeed. The bottom up approach is based on defining the network from known reactions, and if inconsistencies appear, then improvement of the network is needed.

Even with an available genome scale reaction network, what precisely is going on inside the cell and the regulation of the processes are still largely unknown. In systems biology we try to regain a holistic view of the cell, by implementing all available data to describe what is happening (Bull, 2010). This is obviously not uncomplicated, and many papers can be found implementing various 'omics' data into genome scale model networks, and also trying to integrate several different types of 'omics' data together, some of which will be discussed in the following sections (Zhang, et al., 2010) (De Keersmaecker, et al., 2006).

Figure 21 presents a schematic overview of the two main approaches. The black-box modelling is a kinetic descriptive approach that requires little mechanistic understanding of the system to be applied. It is based on simple physiological descriptions, and gives a very good measure of cellular performance; it is one of the ways to describe cellular performance from the experiments so far in the thesis. Alternatively, from the experiments biomass can be harvested under controlled conditions, and further analysed. The sampled biomass, can dependent on sampling technique be used to feed any of the advanced 'omics' analyses. The purpose is to obtain a better description of the internal workings of the cell, including elements such as pathways, the activity of these, and regulation. Finally a predictive model can be generated, which can function as a framework to exhaustively describe cellular performance, as well as predict possible optimisation methods, or the effect of perturbations to the network. The systems biology approach encompasses the integrative use of these methods to gain a holistic description of the cell, and will be the focus of the current chapter.

Related to *Streptomyces lividans* many of the examples given on how the various methods have been used are concerning secondary metabolism, or relative strains. It is therefore important to keep in mind that while experiments trying to map or utilise secondary metabolites can only be related to heterologous protein expression in a limited fashion, then it is also to show how the methods have been developed, and what they are capable of showing. The relevancy of the methods is very much present, despite the examples given will not always seem so.

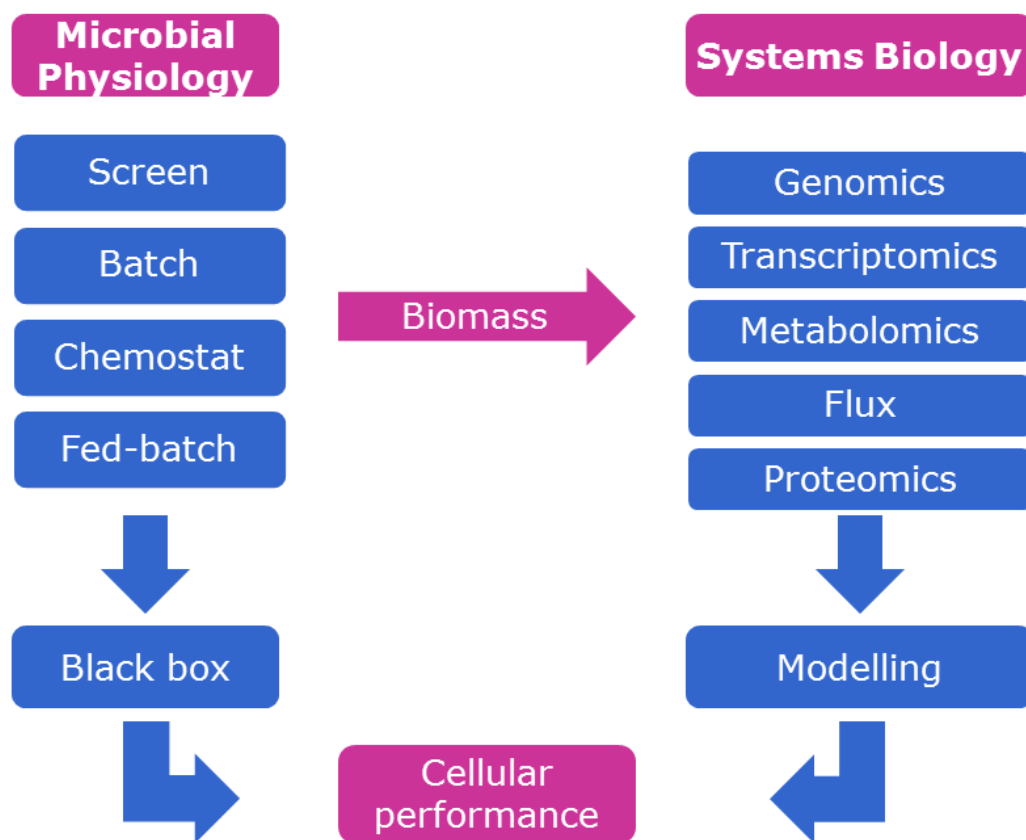


Figure 21 A schematic overview over the preparation of a cell factory, the microbial physiology is divided into the methods usually used to generate biomass. In itself, the physiological approach can feed information into black box modelling, and biomass into the advanced analyses, that can be used to generate more advanced models, that may respond to regulation. Either way, the goal of each of the approaches is to improve cellular performance.

Analytical approaches of looking inside the cells

A number of advanced analytical techniques are available, and have been developed over the last 20 years. These methods represent very diverse disciplines that in the more recent years are being combined under the title of systems biology, which is an attempt to study organisms in a more wholesome view. This holistic approach is not a new thing, rather a return to a methodology that was once common across the fields of science (Bull, 2010). The divergence seems to have been caused by the fact that developing each of the methods required highly specialised skills to be successful. An overview of the methods together with some of the key results, in relation to *Streptomyces* will be provided in the following sections. In relation to systems biology some of the methods have been used very successfully in cooperation, to provide an added level of information, as well as hinting at the underlying regulatory mechanisms (Likic, et al., 2010).

Genetics

Having access to the genomic sequence is not the be-all-end-all approach; a lot of information can be gathered on the physiology of an organism by investing time in the lab growing the strains under various conditions, which can give hints at the capabilities of a strain (Hopwood, 1999) (Paradkar, et al., 2003)

(Bochner, et al., 2001). Genetic studies, as opposed to genomics, predate even the discovery of DNA as hereditary material, and still incredible accurate genetic maps have been designed, without knowing the genomic sequence (Hopwood, 1999) (Kieser, et al., 1992). Hopwood divides the *Streptomyces* genetic research into three eras: *in-vivo*, *in-vitro* and *in-silico*, where the first is the collection and study of naturally occurring, and induced, mutants. The second is the understanding of genetic elements, characterization of the genes, and developing the molecular biology. The last is the rational use of available DNA sequence information, to understand and improve the organism; this does not necessarily entail whole genome sequences. The two major avenues of research of *Streptomyces* that have been followed so far have been divided between the basic physiology including primary metabolism, and secondary metabolism, antibiotics and developmental physiology (Hopwood, 1999). These two avenues are still the dividing line in *Streptomyces* research, and the vast amount of attention is spent on the secondary metabolism fuelled by the economic interest in the metabolites with beneficial properties produced by the Actinomycetes (Paradkar, et al., 2003). The main purpose of introducing the advanced analytical techniques to describe cellular performance in a systems biology perspective is to allow a model of the system to move from a purely descriptive role to a predictive. The wish is to obtain a platform on which it is possible to analyse and interpret high data density 'omics' data, as well as obtain sufficient predictive power to suggest metabolic engineering optimisations with the purpose of improving *S. lividans* as a cell factory. The separate 'omics' analyses, depicted in Figure 21 will be introduced separately in the following sections, with examples of how they can and have been used.

Genomics

Actinomycetes in general have been investigated, mostly for their interesting properties, such as a wide range of bioactive compounds and highly specialised proteins; in a few cases the interest has been caused by the pathogenicity of the strain, e.g. *Streptomyces scabies*, causative agent of potato scab disease. Further, *Streptomyces* strains are often being found inhabiting old industrial areas, such as mines, and these strains have gained resistance towards those environmental protagonists, which makes them further interesting to study (Schmidt, et al., 2009) (Lin, et al., 2011).

With the availability of the genomic sequence, comparative genomic approaches are possible. The published full genome sequences and the sequenced but not published and/or not completely annotated sequences are presented in Table 3. According to the Genome online database (GOLD)², there are 15 completed and published genomes, and a further 22 are considered permanent drafts. In Figure 22 the quality level of the genomes are presented, 6 (7) levels are used, from no level, over first draft (L1) to finished. It is surprising to see 18 are listed as level 6, and only 15 of those are recorded as complete and published. The vast majority are considered incomplete, and have no level. 7 genomes are categorised as "targeted". All except 2 *Streptomyces* genome sequences are publicly available; the two that are not are listed as incomplete, and are *S. venezuelae* and *S. noursei* ATCC 11455, both industrial strains, *S. noursei* is used for producing nystatin and *S. venezuelae* is known as the progenitor of chloramphenicol. It is unclear what exactly a permanent draft is, but in the statistics it groups together with completed sequences, and at least 1 permanent draft has been published (Erxleben, et al., 2011), and is listed as level 5 (where level 6 is complete). Some of the genomes from the Broad institute in Table 3 are considered incomplete, whereas

²GOLD – Genomes online database: <http://www.genomesonline.org/cgi-bin/GOLD/index.cgi> sponsored by US. Dept. of Energy, Joint Genome Institute (DOE-JGI), accessed (06-06-2013)

on Broad's homepage they are listed as drafts; though 1 was complete, and some are registered as permanent drafts. Two of the sequences related to *Streptomyces* are not publicly available. The Broad institute has a large number of draft sequences listed, including that of the *S. lividans* TK24, a derivative industrial strain from the WT *S. lividans* 66; the industrial TK24 strain is without any integrated plasmids (Lewis, et al., 2010).

Table 3 Published and/or finished full genome sequences of *Streptomyces* spp. Annotation of strains can still be progress. n/p is not published (yet) in a journal. All but two should be publicly available.

Published sequences	
Strain	Publication
<i>S. coelicolor</i>	(Bentley, et al., 2002)
<i>S. cattleya</i>	(Barbe, et al., 2011)
<i>S. griseus</i>	(Ohnishi, et al., 2008) (Grubbs, et al., 2011)
<i>S. avermilitis</i>	(Omura, et al., 2001) (Ikeda, et al., 2003)
<i>S. scabies</i>	(Yaxley, 2009) (Bignell, et al., 2010)
<i>S. bingchenggensis</i>	(Wang, et al., 2010)
<i>S. flavogriseus</i>	
<i>S. venezuelae</i> Shinobu 719	(Pullan, et al., 2011)
<i>S. violaceusniger</i>	n/p
<i>S. sp. SirexAA-E</i>	n/p
<i>S. hygroscopicus jinggangensis</i>	n/p
<i>Streptomyces albus</i> J1074	Broad, n/p
<i>Streptomyces</i> sp. PAMC26508	n/p

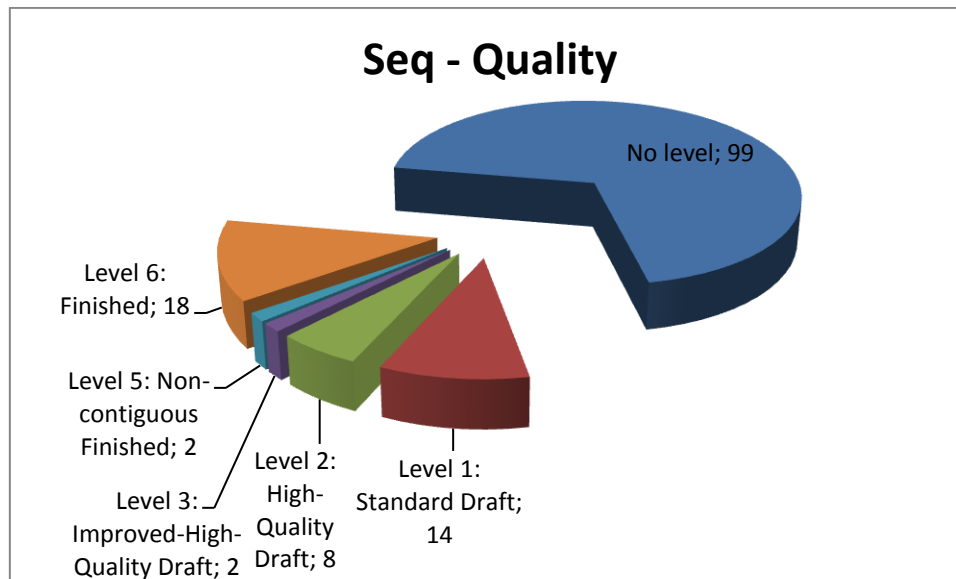


Figure 22 Whole genome sequencing projects overview, for *Streptomyces*, from Genome Online Database (GOLD)

There is a high similarity among the *Streptomyces*, and Actinomycetes in general, which can be used to study related species, without the full sequence information (Jayapal, et al., 2007) (Lewis, et al., 2010) (Paradkar, et al., 2003). The relation is so close, that *S. coelicolor* has been used to immunize a murine infections model against infections from *Mycobacterium spp* (Arzuaga, et al., 2011). A comparative genomic study has also been performed between a number of *Mycobacterium spp*, *Corynebacterium spp*, two *Streptomyces spp*, and a few others, in total 31 organisms (McGuire, et al., 2012). (McGuire, et al., 2012) makes no specific references to *Streptomyces*, as the study concerns pathogenicity, though the ESX transport system is present in *Streptomyces* as well (Chater, et al., 2010), and it does seem that some of the lipid metabolism regulation that is also being investigated shares some regulatory modes between *Streptomyces* and *Mycobacterium*. Despite *Mycobacterium tuberculosis* has a circular genome sized 4.4MB, a substantial amount of synteny is observed within the core region, i.e. excluding the outer 1-2Mb of the *S. coelicolor* genome, which is around 8.6 Mb (Chater & Chandra, 2006). Cross hybridisation of *S. coelicolor* arrays has been successfully utilised in the comparative genomic studies performed between the industrially relevant strains *S. lividans* TK24 and *S. coelicolor* M145, and the WT *S. lividans* 66, using either the spotted array (Jayapal, et al., 2007) or the 104K microarray (Lewis, et al., 2010). It is from these studies evident that the sequence similarity between *S. lividans* and *S. coelicolor* is profound. In relation to this it is not absolutely clear why *S. lividans* produces so sparse amounts of antibiotics, unless specifically prompted, though, differences were found in the *bldB* gene, which is required for sporulation in *S. lividans* and is related to antibiotic production in *S. coelicolor*. The methods used by both (Lewis, et al., 2010) and (Jayapal, et al., 2007) have marked this gene as absent/divergent; and from (Lewis, et al., 2010)'s further investigation, the 5' end of the gene may differ sufficiently that three consecutive probes failed, and marked the gene as absent, although most of the gene is intact. Further, it was discovered that the genes used in the glyoxylate bypass are missing in *S. lividans*, as are totally 9 genes representing core metabolism, including Entner-Doudoroff genes, and breakdown of lactose to glucose and galactose (Lewis, et al., 2010). From the Biolog data presented in chapter 3, it was found that *S. lividans* seem to grow well on lactose as C-source, suggesting either availability of an alternate pathway, or a false classification has taken place. Sequence homology can be advantageous even before performing comparative genomics. Primers from

one strain can be produced, sometimes as degenerate, to prove the presence of genes in other strains, e.g. the discovery of a *Nonomuraea* species using ED pathway to metabolise carbon (Gunnarsson, et al., 2004). In *S. lividans* the ED pathway has not previously been verified, and when a homology search was made (BLAST_n) in the Broad database of the genes of the ED pathway, candidates were found, though they seem to not co-locate on the genome suggesting false positives (Rattleff, et al., 2013b); more information from this study will be presented in the section about Flux analysis, below.

(Zhou, et al., 2012) has made a large scale comparative genomic study between published strains of *Streptomyces*, where more time is spent on categorising the genes, and describing evolution of *Streptomyces* strains. This kind of evolutionary study can be used to help elucidate how the secondary metabolite clusters, and exotic properties have arisen, and possibly help understand the genus. In the study of (Ventura, et al., 2007) the search was broadened, and included different clades, all belonging to the Actinobacterium class, with a heavy representation by the *Actinomycetales* order. The result of the study was the gained broad understanding of the evolution in this order. Many genomes are highly homologous, differing in inversions, and gene duplications. Phylogenetically the Actinobacteria form tight local groups, when based on 16S rRNA sequences. (Ventura, et al., 2007) further discusses the merits of using different methods of ordering the branches of the phylogenetic trees.

Availability of the genomic sequence does give the ability to analyse the capabilities of the cells in a more concerted manner, and maybe discover abilities previously unanticipated e.g. (Bentley, et al., 2002) discovered from the *S. coelicolor* A3(2) genome an unexpected large number of regulatory genes, and several secondary metabolite clusters, as well as three siderophore like proteins hinting at a strong selective pressure for harvesting metal ions. It has later been discovered that it seems to be a general trend that several siderophores are produced by each strain (Chater, et al., 2010). The sequence of *S. avermitilis* was used to deduce the capabilities of the strain (Omura, et al., 2001), and it was found that there was 25 putative secondary metabolite clusters. Even with the genomic sequence available, thorough analysis is needed to verify whether the putative genes are in fact what they are believed to be, and whether, and under which conditions, they are expressed. Experimental conditions impose limitations on the amount of conditions a strain can be subjected to in the lab, which means that it can be difficult to find all types of regulatory sequences. (Studholme, et al., 2004) tested a method to predict regulatory motifs from the known sequence of *S. coelicolor*. A set of putative regulatory sequences were found, and would be suggested to be used in e.g. transcriptomic studies to verify their identities, function and products (Studholme, et al., 2004).

With the sequenced and mostly annotated *S. lividans* TK24 strain, the basis for creating a genome scale model (GSM), as the one proposed by (Lewis, et al., 2010) is present. Following that the 'omics' techniques can be used to strengthen and verify the model, gaining not only a thorough physiological knowledge, but understanding of the regulatory processes at work, which could prove to be very valuable, since the strain is capable of producing Actinorhodin either when prompted by the heterologous expression of a transcriptional activator (e.g. act II ORF 4), or under conditions of very low growth rates (as in the chemostats with D less than 0.035h^{-1} , see chapter 5. This can be utilised to prevent the production of antibiotics, and secondary metabolites in general, without hampering the strain, or preventing sporulation.

Transcriptome

The analysis of all transcribed genes, the transcriptome, requires the sequence knowledge of all, or at least the major part of, genes. There are two major technologies when it comes to transcriptome analysis, hybridization- or sequence-based approaches. (Jayapal, et al., 2007) describes the use of spotted arrays, that have higher flexibility in their design, but eventually less fidelity, as one probe is used per gene. (Lewis, et al., 2010) presents the use of high fidelity 104K arrays, which means they carry 104,000 probes, thus several probes per gene; a number of *Streptomyces* have in the range 7,500 – 10,000 genes (ORF's) (Zhou, et al., 2012), allowing both several probes per gene, thus ensuring redundancy, as well as introducing the ability to test the intergenic sequences for possibly regulatory RNA. An emerging technology is gaining foothold, RNAseq (Wang 2009), since it can be used without knowing the sequence, as it is based on sequencing approach, and therefore can also be used to determine regulatory transcripts (Marioni, et al., 2008). Alternatively it is possible to monitor only a subset of genes using a reduced array, or qPCR (Marioni, et al., 2008), as in the study by (Facey, et al., 2013), where only specific osmotic shock genes were observed in several strains and transcription levels were compared. This way only the sequence of the required genes are needed. On the other hand, in strains, or genera, with a high sequence homology the array from one sequenced strain can be used (although with care) on an unsequenced relative. This has been reported by (Jayapal, et al., 2007), and later (Lewis, et al., 2010), where comparative genomics have been based on a *S. coelicolor* array used for *S. lividans*. Depending on the arrays used the level of detail and resolution varies (Lewis, et al., 2010). An RNA-seq approach was used to investigate the copper sensing in *S. lividans*, though the results seem pleiotropic, or possibly copper toxicity distort the results, since 400 genes show higher expression in the strain containing the regulator knock-out, and none of these genes contain the expected regulator recognition sequence (Dwarakanath, et al., 2012). Using one of the older DNA microarrays platforms, (Huang, et al., 2005) discovered that regulation of secondary metabolism in *S. coelicolor* seems to be an interconnected regulatory network rather than the hierarchical network, which had been the precious dogma. The individual regulators which, previously were thought to be pathway specific seems to be able to extend a certain degree of control on parallel pathways and/or secondary metabolite cluster; it was even observed that there is an ability to modulate some of the more global regulators (Huang, et al., 2005). In a different case, a parallel transcriptomic and proteomic experiment was set up to test the effect of phosphate regulation in *S. coelicolor* (Martín, et al., 2012). There seemed to be a good correlation between the tested 'omics' methods, and a remodelling of the central carbon metabolism was found, possibly in response to a redox imbalance caused by the phosphate depletion (Martín, et al., 2012).

As mentioned previously, the amount of experiments focussed on the primary metabolism of streptomycetes, especially with focus on developmental biology, is quite limited. A study of the pre-sporulation mycelium of *S. coelicolor* growing on solid medium was performed by separating the cells at different developmental stages using protoplastation, and performing transcriptomic analyses on the separate developmental stages (Yagüe, et al., 2013). It was found that the cellular program started, even while the substrate mycelium still exists is consistent throughout all stages of MII (from mycelium, over aerial mycelium to spores). It seems the program entails programmable cell death of the mononucleated mycelium and simultaneous generation of multinucleated mycelium, separation of nuclei and lastly sporulation. Interestingly it seems as though one of the genes found to be differentially regulated was *whiJ*, which seem to act like a sporulation inhibitor (Yagüe, et al., 2013). These studies seem to agree with the

developmental biology as described by (Flärdh & Buttner, 2009). The group has also studied the developmental biology through a proteomic approach (Manteca, et al., 2010).

Transcriptome analysis is valuable in that it gives an indication of what genes are used in a given set of conditions, of even more useful in comparing sets of conditions, where the change in transcription can indicate the use of a gene. This methodology is useful in determining the functional genomics; the association of genes to their function. As was the case of (Borodina, et al., 2005), creating a genome scale model of *S. coelicolor* entails it is also incredibly useful in discovering gene products with regulatory properties. The level of transcription, while informative, is not necessarily the same as a higher flux, or more enzyme product, due to other levels on regulation. It is a fantastically useful tool, in combination with other of the 'omics' analyses.

GSM

The purpose of creating a genome scale model is to perform investigative analyses, such as to run simulations of the cell under different conditions, evaluate the possible effect of genetic manipulation, search for metabolic engineering targets, or help interpret experimental data (Durot, et al., 2009). It is possible to simulate cellular growth based on these models, and determine the effect on growth rate a perturbation might have. All these purposes can be summarised under the title *in silico* biology. If the model is comprehensive, and the predictive power is sufficient, then multiple conditions can be tested, before verification is required in the lab. From the outset, a genome scale model can be developed as either a top-down, or bottom-up approach (Palsson, 2002). The top down approach is aiming at creating a network from observable interactions. The bottom-up is creating a network from comprehensive knowledge of known interactions, typically assembled from literature. Ideally the top-down method should be resolved on bottom-up foundation, to reduce computational complexity (Oberhardt, et al., 2009), and obtain the most comprehensive model (De Keersmaecker, et al., 2006). In either case, the Top down approach alone is so data and computationally demanding, that it is almost unviable to not support it by bottom-up information, though as a supplement it can likely reveal interactions that have yet to be verified otherwise. Many types of software have been developed to support the bottom up network reconstruction i.e. to help to interrogate databases, and organise the resultant data into reaction networks. However, eventually all reactions have to be manually curated (Feist, et al., 2009) because the annotations are often lacking in detail, such as "hypothetical protein", or are downright false, based on simple BLAST homology search, and therefore can point to a different protein, or class of protein (Borodina & Nielsen, 2005). Using the proven literature as the scaffold, and the supplementing with knowledge gained from the sequence should be the method of choice.

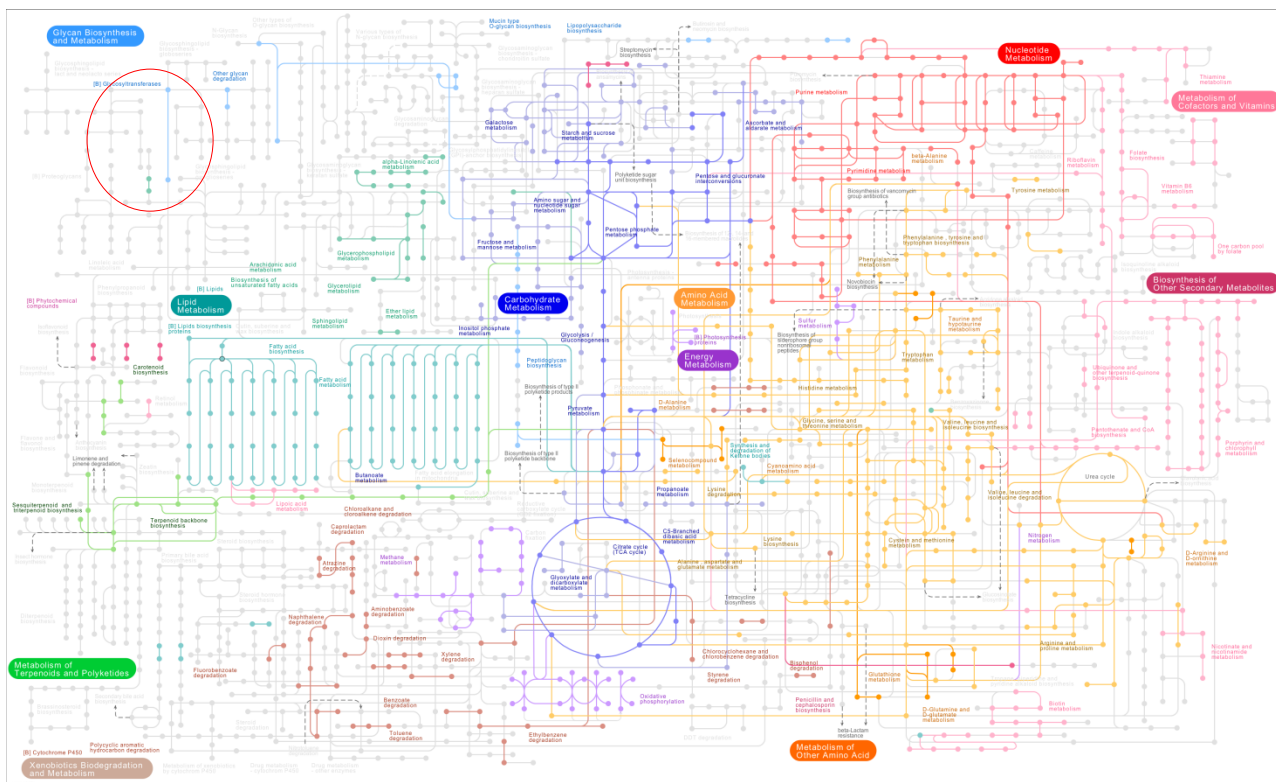


Figure 23 All reactions in KEGG, that are annotated for *Streptomyces coelicolor*

One of the large databases of annotated reaction available in various organisms (including some *Streptomyces*) is KEGG, in Figure 23 the reaction network is presented, with all annotated reactions fully coloured, and the weakly coloured are possible reactions. It should in this figure be noted that some “islands” can be observed, metabolites or short reaction pathways without seeming connection to the central carbon metabolism, as those marked with a red circle in the top right corner of Figure 23. This is a typical result, and will require gap-filling in the model produced, either from homology search from other *Streptomyces*, or from computational inference, where the dead-end or orphan metabolites are attempted connected to the network from a repository of reaction, using the GapFill method (Durot, et al., 2009). When a genome wide network has been produced, the initial test is to see if all components required for biomass can be produced, and secondly compare physiological data to the results from the model, and determine if reactions are missing (Reed, et al., 2006). Alternatively, phenotype arrays such as those produced by Biolog can produce hints for missing reactions, or metabolic capabilities. The network does not require the knowledge of the genomic sequence, though this knowledge will allow expanding the network to include even pathways that haven’t been (well) characterised in the specific organism yet. Further, with sequence data, and transcriptional data, regulation can be more thoroughly investigated.

The network created is the basis for investigating cellular performance deeper. It is the framework around which intracellular modelling can be performed. The models can be used to investigate the network topology and function, estimate flux distributions, integrate high density data such as transcriptome, and guide metabolic engineering efforts. The first models for flux analyses were created from the bottom-up; some even before the genomic sequences of the strains were available (Nowruzi, et al., 2011). Such models have been created as a reduced network due to computational simplicity. These types of models can be

found for most organisms, and as such (Borodina, et al., 2005) has created a model for *S. coelicolor*, which was to expand on the simple models used by (Rossa, et al., 2002) and (Bruheim, et al., 2002).

The model of *S. coelicolor* presented by (Borodina, et al., 2005), consisted of 700 unique reactions, and a total of 971 reaction, including iso-enzymes. The reactions included in the model represent 13% of the estimated ORFs in *S. coelicolor*. Due to the vast homology between *Streptomyces*, a model like this can be used on other members, with minimal adaptation. An example of this was presented by (Lewis, et al., 2010) in a cross hybridization study between *S. lividans* and *S. coelicolor*, where a number of genes were found to be divergent/absent in *S. lividans*. The genes were identified by genome wide hybridisation using a 104k microarray (Lewis, et al., 2010), but substantiated the findings performed on a spotted array (Jayapal, et al., 2007). Of all the genes found absent/divergent 9 were related to the central carbon metabolism. The authors removed the reactions from the genome scale model produced by (Borodina, et al., 2005), and tested whether loss of these genes would result in a likely viable cell, which they would, indicating that these absent/divergent classifications are plausible. The models that have been developed previously have mainly focused on steady state conditions, due to the strength of the linear description of the model (Durot, et al., 2009). The limitation this imposes is that when secondary metabolism is studied, the switch from the exponential growing stage to the production of secondary metabolites in the stationary phase has had to be modelled as two separate events. (Alam, et al., 2010) has succeeded in creating a model that can model the transition to the production stage. The authors say that including sufficient data allows modelling the switch, and they included more than 25 transcriptomic samples over the time course of the experiment. The final model was compared to other published models, and was found to compare well to the model produced by (Borodina, et al., 2005), although expanded on the part of the secondary metabolism, so the total model includes 1015 reactions representing 789 genes. The model by (Alam, et al., 2010) uses both flux balancing analysis (FBA) (Price, et al., 2004), to create a plausible solution space, and included a dynamically changing input function to enable using the time series data represented by the large number of transcriptomic samples.

(De Keersmaecker, et al., 2006) and (Zhang, et al., 2010) present in their reviews the current state of integration of various types of 'omics' datasets, and describes some of the difficulties, as well as promise, this presents. 'Omics' data is in its nature noisy (see e.g. (Alam, et al., 2010)), but when using 'omics' data from several different methods, they can supplement each other, and reduce the impact of the noise in each dataset. Further, several levels of 'omics' data integration also carries the promise of discovery of regulatory systems that have previously eluded discovery. Integrating several complex datasets or using these as a basis for generating genome scale networks may result in falsely estimating the impact of a change in the network; therefore, when studying the network, the topology as well as the function of the network has to be considered (Mahadevan & Palsson, 2005). Network analysis has resulted in determining highly connected nodes (metabolites) as crucial, but many linear pathways are known, and knocking any gene out in these will result in the functional loss of the final product, not only the highly connected precursor.

While some of the modelling procedures have been developed specifically to focus on secondary metabolism, especially the metabolic switch as studied by (Alam, et al., 2010), the methods can be used as part of a primary metabolism toolbox. Knowing, what and how secondary metabolism is triggered, can offer targets for engineering for reducing the impact, while not losing the ability to sporulate. Sporulation is

a useful feature to possess when storing the strains, and creating cell banks. Many of the regulators have historically been identified by the loss of the ability to grow aerial hyphae (the *Δbld* class of mutants), or the loss of sporulation (the *Δwhi* class of mutants). Relying on the models, targets could be sought with less impact in the cellular metabolism since e.g. *ΔbldA* (encoding the tRNA that recognises the rare UUA codon, and has been connected to the secondary metabolism) had effects on the primary metabolism (Hesketh, et al., 2007).

Proteome

Like the other 'omics', the aim of proteomics is to give a picture of all present proteins in a given sample. The motivation is that the proteins are the workhorses of the cell, and reflect what the cell does to try and cope with an external stimulus. To go from stimulus to protein, in a measurable quantity, does take some time, but it is none the less a more stable sample. The turnover time of proteins is a lot longer than for metabolites, and as such makes the analysis more robust. Further, when analysing proteins a further level of regulation can be observed. Proteins have their activity modified by physical change of the protein, e.g. phosphorylation or methylation. Therefore protein regulation can be directly observed. The conventional methods of separating proteins either by mass/charge or isoelectric point yields an insufficient separation, as several proteins will co-elute. 2D-electrophoresis has been the standard operating procedure since the 70's, because a good separation is achieved, even between variants of the same protein with different regulatory modifications, separation can be achieved. Still, 2D gel electrophoresis is being blamed for only resolving in the 100's of proteins, which is a small subset of the total cellular capability. The more recent use of liquid chromatography coupled to tandem MS, have allowed detection of 1000's of proteins. Even the low abundance proteins have become accessible (Zhang, et al., 2010). Unfortunately only 20-40 % of the total proteome can be detected, and so far only qualitatively (Zhang, et al., 2010).

S. lividans possesses 4 extracellular signal peptidases (*sip* family), where organisms can be found harbouring anything from 1 SP, which is essential, to 7 as seen in *Bacillus subtilis* (Escutia, et al., 2006). Signal peptidases are responsible for processing the proteins that are exported by the classical SEC transport system. 3 of the signal peptidases are located in the same operon, and one is located separately on the genome. It seems that *sipY* is the major peptidase, however loss of the peptidase can be complemented by overexpression of either of the other 3 (Escutia, et al., 2006). In this work the initial estimation of the extracellular proteome was pictured on strains growing in casamino acid supplemented minimal medium. A number of proteins were identified in the extracellular medium, including some previously annotated as hypothetical proteins, opening for the possibility of ascribing functionality to them. The work is by the authors own statement not an exhaustive study. An alternative is to overexpress a (putative) regulatory gene, and using the proteomic snapshot to try and identify differentially, or newly, expressed proteins, on par with, or as supplement to, a transcriptomic analysis (Kim, et al., 2005) (Choi, et al., 2010). (Li, et al., 2006) and (Hesketh, et al., 2007) have studied the effect of deleting the *bldA* gene in *S. coelicolor* using a combination of transcriptomics and proteomics analyses. A large number of genes were identified as affected by the deletion of the tRNA recognising the TTA (UUA) codon, even gene products from the primary metabolism. A surprisingly small number of the affected (less transcribed) genes had the TTA codon as part of their gene, or transcription initiation sequence. One of them were SCO6717, annotated as "putative acyl-[acyl-carrier protein] desaturase" and a target for TTA regulation (the TTA codon was upstream of the start codon) (Hesketh, et al., 2007). Further, (Hesketh, et al., 2007) mentions

that most of the TTA regulated proteins may be related to surface growth, or has ecologically adaptive functions; therefore these proteins will likely not be identified in liquid culture.

Many of the advanced analysis methods have so far been used by groups specialising in the analytical technique, which results in samples being taken in a biological unstructured fashion. The experimental setup does obviously influence the results, and sampling different strains after a specified amount of time, e.g. 8 days, irrespective of growth phase will result in unwanted effects. This is commonly observed in literature, that sampling is undertaken either regardless of growth phase, or not normalised across strains to a feature e.g. biomass, which tells whether an e.g. increased production is caused by better productivity or higher biomass.

A number of tags and isotopes have been developed to give the quantitative view of proteomics. One of these is the iTRAQ, which has been used for both relative and absolute quantification in proteomics. This method has been used to describe the developmental biology of *S. coelicolor*, as well as suggest proteins that change subcellular localisation across developmental stages (Manteca, et al., 2010). Another study (van Veluw, et al., 2012) found that in rich media, two distinct populations of pellets sizes co-exist in streptomycetes, with the smaller of the populations being almost independent of time (265µm), and the larger varying from (300-500µm). Further, this study also found that of the proteins more abundant in the large pelleted population was stress related enzymes, possibly due to mass transfer e.g. oxygen; but also antibiotic synthesis proteins (Calcium dependent antibiotic CDA) were overrepresented in the large population. The smaller population had an overrepresentation of DNA handling proteins, suggesting that the smaller sized population is growing rapidly. A previous study (van Wezel, et al., 2006) has suggested that the smaller sized population has the highest production of proteins, and a preference for septation and fragmentation can be obtained by overexpressing *SsgA*,

A subtype of proteomics is the interactome, defining and describing interaction complexes, such as those of the primary carbon pathways, where the proteins assemble into closely coordinated protein complexes. As such, interactomics is a valid description of all interactions, including genetic and other molecular interactions. In proteomics, interactomics can be used to identify proteins of unknown function, by identifying its interaction partners, and looking at homology.

Proteomics, besides enabling identification of regulatory mechanisms, is extremely useful in supplying a snapshot of the actual capabilities of the strain at a given time, or under a specific set of conditions. Proteomic methodologies have been used since the 70's, and have steadily been developed. Like other of the 'omics' technologies one of the hurdles is that it is difficult to indisputably identify all elements present in a sample, and since proteomics relies on partially degraded samples there is a risk of detecting false positives. The field is constantly being developed, and large online databases exist, which can be used for homology searching, or generating *in-silico* mass-fingerprints for comparison. Proteomics methods have been used in trying to identify the effect of regulatory molecules, in various species of *Streptomyces*.

Metabolome

Metabolome analysis is a much discussed field, and being able to measure the identities and pool sizes of compounds/metabolites in the cell at any given moment would yield a tremendous amount of information, as well as integrate with other 'omics' analyses. The scourge of the analytical chemists is that it will likely never be a complete image, due to the very diverse chemical properties of the compounds, and the

available technology for detection. There is no “black-magic-box-machine-from-CSI”, which allows to measure all metabolites in one go (Dunn, 2008). With proper sampling technique, and with sufficient amount of sample available, it is becoming possible to analyse still larger groups of metabolites, such as phosphorylated compounds, energy metabolites, sugars, amino acids, organic acids etc. One of the most challenging aspects of metabolomics is that due to the very short turnover time, in the less than seconds range, and the small pool sizes, the quenching is of paramount importance, when considering metabolomics (Wittmann, et al., 2004). Before anything can be studied, the sample has to be withdrawn from the culture, in a representative manner, and the sampling should be quick enough that the sample does not change significantly in the process. Quenching is the process of halting the metabolism, and methods rely on either very high or very low temperatures, coupled to a shock like a shift in pH (Villas-Bôas, et al., 2005).

One of the problems with quenching aside, from the fact that the sampling process has to be fast enough to give a proper image of metabolism, is the loss of intracellular metabolites, resulting in the mix of intracellular and extracellular metabolites by organic solvent disruption of the cell envelope (Villas-Bôas, et al., 2005). For quenching the most often used method relies on methanol, which is a strong solvent, and various amounts of leaking have been observed between different organisms and different quenching solutions (Bolten, et al., 2007). Leaking is undesirable, since it causes the underestimation of pool sizes and it is useful to be able to separate the intracellular and extracellular environment when describing metabolism (Villas-Bôas, et al., 2005) (Villas-Bôas & Bruheim, 2007). This has prompted a search for other methods/solvents, some of the tested methods has been pure cold methanol, methanol:water 60:40, methanol:buffer 60:40, cold glycerol:saline 1:1, liquid nitrogen, cold 10M formic acid (Villas-Bôas, et al., 2005) (Villas-Bôas, et al., 2006) (Villas-Bôas & Bruheim, 2007). Cold, in relation to metabolomics usually means temperatures lower than -30°C, except glycerol:saline, where it is -20°C since lower temperatures causes the solution to freeze. (Dietmair, et al., 2010) worked with mammalian cells, and found least leakage, as well as an efficient halt of metabolism, proved by estimating the ATP to ADP conversion, when using 0.9% saline at 0.5°C. Mammalian cells are physiologically different, and (Bolten, et al., 2007) found that low ionic strength wash solutions, including 0.9% saline, resulted in leakage in bacteria, while an optimum was found using 2.6% saline, and not a significant reduction, for higher concentrations up to 10% saline. This does though indicate that sampling should be thoroughly tested and that the very harsh methods reported in literature may not be needed (Dietmair, et al., 2010) (Sellick, et al., 2009).

The next step is the extraction procedure. Already a selection of available metabolites has been made by the choice of quenching method (Fiehn, et al., 2006). This further complicates matters, and may result in having to compromise on a less efficient method (Maharjan & Ferenci, 2003). Extraction methods include disrupting the cell membrane, which can be done by freeze/thaw cycles in 100% methanol, which would support that methanol damages the cell membrane, boiling ethanol or acidic/alkaline extraction (Kassama, et al., 2010). Depending on what is to be investigated, the extraction method has a great impact; solvent based extractions have yielded good results with amino acids, and central carbon metabolism related compounds, though for energy metabolites or tri-phosphates acetonitrile seems to be better (Rabinowitz & Kimball, 2007). In a study by (Dietmair, et al., 2010), acetonitrile also proved to be one of the best extraction methods, for all compounds, amino acids, sugar, organic acids, NAD(P), nucleotide- tri-, di-, and mono-phosphates for mammalian cells.

Due to the complications mentioned above, metabolomics is generally divided in two main approaches. One option is the targeted search for specific (classes of) metabolites, the other option is a wide screen of all detectable metabolites present e.g. metabolic finger/foot-printing, for investigating as many metabolites as possible either intra- or extracellular (D'Huys, et al., 2011) (Kassama, et al., 2010) (Villas-Bôas, et al., 2006). High numbers of compounds are being reported one such report is the identification of 198 out of 453 metabolites (Likic, et al., 2010), though the ability to identify more metabolites will likely increase over time, and with development of both technology and methodology. And one thing is detection of the metabolites, in one or all of the samples; another entirely different matter is quantifying the metabolite, which usually yields fewer metabolites. The metabolic footprint, or the exometabolome, was used by (D'Huys, et al., 2011) to characterise the amino acid uptake in a complex growth medium. They found a large amount of alanine and pyruvate to be produced, possibly as part of overflow metabolism, and suggesting further investigation, as well as a metabolic engineering target. In a qualitative study by (Kol, et al., 2010), osmotic shock from salt stress was studied using a metabolic fingerprint approach. As many metabolites as possible was searched for, and based on database homology, 229 metabolites received a putative identity. Further, grouping by hierarchical clustering found that 143 metabolites sorted into 4 types of clusters, where 33 metabolites were consistently more abundant under salt stress, 19 less abundant, and 91 metabolites showed ambiguous, or dynamic, response.

A number of promising techniques are being developed along the way, and some aim at linking 'orphan' metabolites in a functional genomics manner, and this way an unknown pathway was discovered in *E. coli* (Saito, et al., 2010). Orphan metabolites are those that can be detected/identified, but so far no plausible accounts have been made of the pathways or gene-products used to create these metabolites. This establishes the possibility to improve annotation of the genome, and extend any existing metabolic model.

Metabolomics can also be used for enzymatic characterization in an *in-vitro* setting, by monitoring the changes in the metabolite pool *in-situ*, as well as verification of possible pathways predicted *in-silico*, by unequivocally identifying the products and intermediates of the pathway (Saito, et al., 2010). Using a strong separation technique, followed by different MS detection methods, allow for accurate mass determination (indicative of the molecular formula), fragmentation patterns (indicative of molecular structure) (Dunn, 2008), thus enabling detection of known metabolites, or identification of new/unknown metabolites.

Metabolomics, despite its limitations, can yield a great deal of information in terms of identification and verification of pathways, as well as the extremely powerful coupling to genome scale, or metabolic flux models, where the data can be used to constrain the models, thereby improving their predictive capacity.

Flux

One of the basal steps to elucidate the internal workings of a cell is found using metabolic flux analysis (MFA), which is aimed at describing the activity of the reaction network, under a given set of conditions. For instance, MFA can be used to compare producing and non-producing conditions, to investigate if there are some obvious limitations/bottle-necks in precursor supply that could be targeted for metabolic engineering (Nielsen, 1998). While MFA does not require knowledge of the genomic sequence *per se*, it does require a basic knowledge of the metabolism, the stoichiometric relationship of the network (Daae & Ison, 1999), to obtain a somewhat useful reaction network. This could be the core of the central carbon metabolism, TCA, ED, EMP, PP pathways, and subsequently lump all reactions that eventually lead to biomass accumulation. This would constitute a minimal reaction network; alternatively, if the genome is

sequenced and annotated, a genome scale reaction network can be made from the information in genome and annotation databases. MFA relies heavily on $\text{mass}_{\text{in}} = \text{mass}_{\text{out}} + \text{accumulation}$, which means that all extracellular metabolites should be measured, and uptake/production rates calculated, in order to constrain the solution space of the model. Many of the flux models that are in use today are based on solving the network as a linear problem (Gombert, et al., 2001) (Christensen & Nielsen, 2000) (Christensen & Nielsen, 2000) (Daae & Ison, 1999) (Sauer, 2006), due to the computational simplicity of such system, and requiring only a steady state condition (Kauffman, et al., 2003). One of the problems in MFA is that the network is likely underdetermined; (Daae & Ison, 1999) attempted to overcome this by creating the *S. lividans* network, and balancing it to prevent dependent reactions, and therefore some reactions had to be omitted; the network of (Daae & Ison, 1999) was 53 metabolites and 57 reactions. (Borodina, et al., 2005) created a genome scale network for *S. coelicolor*, based on genome annotation, containing 971 reaction, including transport processes, and 500 metabolites. Based on the sensitivity analysis by (Daae & Ison, 1999), a relatively big change (20%) in monomeric demands on biomass formation, did not significantly affect the overall flux, and therefore *E. coli* biomass requirements were assumed to be acceptable for use in *Streptomyces*; whether this is a feature of the limited network presented by (Daae & Ison, 1999), is unknown, but the assumption propagated to the model presented by (Borodina, et al., 2005).

MFA relies on measured metabolites and rates to constrain the model to an acceptable solution space, though all intracellular reactions are unconstrained, unless further measurements take place. It is possible to manually include information from the literature for each specific reaction about the highest possible rates in either direction, despite this; the model may not produce a single optimal solution (Price, et al., 2004). Another way to restrict the network is via the use of Flux Balance Analysis (FBA), which is a mathematical method of determining the allowable fluxes through each reaction, and thereby creating a possible flux solution space (Varma & Palsson, 1994). One way of introducing information that can help constrain the model, is to make measurements relating to the intracellular milieu, this could be coupling the metabolome to the analysis, which would constrain the model very efficiently, assuming the measurements are reliable (Lee, et al., 2006). Another option is to feed the organisms with isotope labelled substrates (Christensen, et al., 2002) (Christensen & Nielsen, 2000), followed by hydrolysis of the biomass, to investigate the label incorporation, which gives very reliable information on how carbon exchange has taken place in the reaction network. Using labelled substrates (e.g. $^{13}\text{C}_1$ -glucose) allows to identify active pathways, e.g. in the central carbon metabolism, the labelling pattern of pyruvate tells exactly how active ED, EMP and PP pathways are, since PP results in unlabelled pyruvate, due to decarboxylation of C_1 , ED results in 50% labelling in C_1 of pyruvate, and EMP results in 50% labelling in C_3 of pyruvate. The reason for 50%, and not 100%, is that 1 $^{13}\text{C}_1$ -labelled glucose molecule will be broken into 2 pyruvate molecules. This is exactly how we found which actinomycetes have an active ED-pathway, without knowing the sequence of many of them (Rattleff, et al., 2013b), the draft of the paper is found in appendix C. Other labelled substrates can be used, e.g. $\text{U-}^{13}\text{C}$ -glucose, which can be used to supplement $^{13}\text{C}_1$ -glucose, or used when several substrates are being used (Fischer, et al., 2004).

Many optimisation possibilities are available when considering flux analyses. It could be the calculation of the optimal growth medium, minimisation of by-product formation, identifying metabolic engineering targets for optimal product formation (Varma & Palsson, 1994). (Bushell, et al., 2006) has used a variant of FBA, flux variability analysis, to define medium composition for a *S. coelicolor* *zwf* mutant. Further, integration with e.g. metabolomics seems very promising, and would allow reducing the degrees of

freedom in the flux system (Lee, et al., 2006). FBA can be used to determine the possible solution space for fluxes, and when investigating knock-outs a variant of FBA has been developed called minimisation of metabolic adjustment (MOMA), which aims at most precisely describing how the mutant strain acts in relation to its wildtype counterpart, since simple optimization has shown to be imprecise (Lee, et al., 2006) (Price, et al., 2004). Further, flux models have begun to include Boolean rules to describe regulatory control in the flux system (Lee, et al., 2006) (Kauffman, et al., 2003).

MFA can be used to understand how the cell utilises various medium components, and in extension thereof increase the understanding to improve certain traits of the cell. In the genome scale flux model presented by (Borodina, et al., 2005) the co-factor requirements are included, which can help predict imbalances when producing a specific product. As was the case for *S. coelicolor* and *S. lividans* producing actinorhodin and undecylprodigiosin, (Borodina, et al., 2008) found that deleting the phosphofructokinase (*pfk2*) in *S. coelicolor*, and forcing flux through pentose phosphate pathway increased the antibiotic production; whereas (Butler, et al., 2002) found that a reduced PP-flux, in *S. lividans*, gave higher production. (Butler, et al., 2002) limited the flux through the PP pathway by deleting one of two *zwf* genes, or the *devB* gene encoding some of the early steps in the PP-pathway. (Butler, et al., 2002)'s results were supported by (Rossa, et al., 2002), who ran flux analysis on chemostat grown cells. It is surprising that two so closely related organism react so differently, and (Borodina, et al., 2008), does speculate that overflow metabolism may be involved, and further concludes that the central carbon metabolism is so tightly regulated that overexpression of enzymes may not be enough to counter the effect of flux controlling steps, that should rather be deleted, if deregulation is sought.

An area that has been of much interest concerning flux and *Streptomyces*, is the connection between the primary and secondary metabolism. Many of the antibiotics, or other secondary metabolites, are produced from precursors drained from the primary metabolism, and therefore has this specific utilisation attracted some attention. Some of the studies have already been mentioned (Bushell, et al., 2006) (Borodina, et al., 2008) (Butler, et al., 2002) and (Rossa, et al., 2002). When the drain from primary metabolism is amino acids, or intermediates in amino acid metabolism, it seems there is a positive correlation between supplying the needed precursors in the medium, and product formation (Bushell, et al., 2006). A study by (Kiviharju, et al., 2007) indicated that for the specific strain they were studying, it seemed that a shift from growth to product formation could be obtained best by perturbation of the pH. Further, it was found that phosphate limitation is very effective for initiating secondary metabolism, which by now has become a well-known fact (Kirk, et al., 2000). It seems that there is an inverse relationship between growth rate and flux through PP, TCA and shikimate pathways (Kiviharju, et al., 2007). For TCA a similar relationship was observed in *S. noursei* by (Jonsbu, et al., 2001). Whether these relationships are significant in terms of product formation, or if they are a feature of the strains specifically is not known (Rattleff, et al., 2013b), but it is interesting that the same optimisation criterion, namely antibiotic production, can be achieved by opposite means in two closely related strains, *S. lividans* and *S. coelicolor* relating PP-flux (Borodina, et al., 2008) (Bushell, et al., 2006). According to (Bruheim, et al., 2002) calcium seems to be rather important for obtaining high biomass concentrations, and seems to have an inverse relationship with actinorhodin, suggesting that calcium in excess should be good for protein production, assuming no other untoward regulation occur. (Ryu, et al., 2006) found that deletion of *pgm* (phosphoglucomutase) was expected to yield higher production of actinorhodin, because a smaller amount of carbon was expected to be sourced into glycogen. The findings were the opposite, suggesting both that *pgm* seems to be equally related in

build-up and breakdown of glycogen, as well as the understanding at present about glycogen metabolism needs to be elucidated. Overexpression of the ACCase (Acetyl Co-enzyme A carboxylase) gave very high improvements in the production antibiotics (actinorhodin and undecylprodigiosin).

Two studies have been found, specifically addressing flux analysis of protein production in *S. lividans* (D'Huys, et al., 2012) and (Nowruzi, et al., 2011). The paper by (D'Huys, et al., 2012) is an extension, of (D'Huys, et al., 2011), where the extracellular metabolome was followed to determine amino acid uptake profiles, for a heterologous protein producing strain of *S. lividans*. The data has then subsequently been used to define the bounds of a flux analysis, in pseudo steady state, in late exponential growth. The findings of (D'Huys, et al., 2012) are that biomass formation is not a suitable optimisation goal, and a second goal has to be added for the predictions to fit the data. The secondary goal was found to give the best agreement, when using optimal ATP yield, as secondary objective. Further, little flux was found to go through PP-pathway, and NADH generation seems to happen in TCA (D'Huys, et al., 2012). The results of (D'Huys, et al., 2012) are obtained purely as flux analysis, without isotope labelled substrates, there is therefore room for further elucidation of the findings. It is however an excellent example of integrating metabolomics and flux analysis. (Nowruzi, et al., 2011) uses a relatively simple model, of the central carbon metabolism, and finds that the growth rate limitation is not carbon but nitrogen uptake, either amino acids, in the early growth phase (E1), or ammonium in the 'late' growth phase (E2). Interestingly, the growth rates found by (Nowruzi, et al., 2011) are all below 0.08h^{-1} , which is a lot less than what I have observed during my experiments. As far as can be determined, the largest difference between the medium composition of (Nowruzi, et al., 2011) (Nowruzi, et al., 2008) and mine, is that I add a vitamin solution (1 mL/L, see chapter 3), which may relieve further growth limitation.

Flux analysis, with its various derivations, has been exceedingly helpful in determining the intracellular division of carbon, in the *Streptomyces*. Successful studies have both improved growth and product formation (whether heterologous protein or antibiotic), medium compositions have been optimised and links between the primary and secondary metabolism has been established. With the extension of the other 'omics' analyses there is great hope that the various flux analyses can become even better predictive tools.

Works Cited

- Alam, M. T. et al., 2010. Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics*, Årgang 11, p. 202.
- Arzuaga, N. O. et al., 2011. The use of *Streptomyces* for immunization against mycobacterial infections. *Hum Vaccin*, Sep, 7(9), pp. 934-940.
- Barbe, V. et al., 2011. Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites. *J Bacteriol*, Sep, 193(18), pp. 5055-5056.
- Bentley, S. D. et al., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, May, 417(6885), pp. 141-147.
- Bignell, D. R. D. et al., 2010. *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Mol Plant Microbe Interact*, Feb, 23(2), pp. 161-175.
- Bochner, B. R., Gadzinski, P. & Panomitros, E., 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res*, Jul, 11(7), pp. 1246-1255.
- Bolten, C. J. et al., 2007. Sampling for metabolome analysis of microorganisms. *Anal Chem*, May, 79(10), pp. 3843-3849.
- Borodina, I., Krabben, P. & Nielsen, J., 2005. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res*, Jun, 15(6), pp. 820-829.
- Borodina, I. & Nielsen, J., 2005. From genomes to in silico cells via metabolic networks. *Curr Opin Biotechnol*, Jun, 16(3), pp. 350-355.
- Borodina, I., Schöller, C., Eliasson, A. & Nielsen, J., 2005. Metabolic network analysis of *Streptomyces tenebrarius*, a *Streptomyces* species with an active entner-doudoroff pathway. *Appl Environ Microbiol*, May, 71(5), pp. 2294-2302.
- Borodina, I. et al., 2008. Antibiotic overproduction in *Streptomyces coelicolor* A3 2 mediated by phosphofructokinase deletion. *J Biol Chem*, Sep, 283(37), pp. 25186-25199.
- Bruheim, P., Butler, M. & Ellingsen, T. E., 2002. A theoretical analysis of the biosynthesis of actinorhodin in a hyper-producing *Streptomyces lividans* strain cultivated on various carbon sources. *Appl Microbiol Biotechnol*, May, 58(6), pp. 735-742.
- Bruheim, P. et al., 2002. High-yield actinorhodin production in fed-batch culture by a *Streptomyces lividans* strain overexpressing the pathway-specific activator gene actII-ORF4. *J Ind Microbiol Biotechnol*, Feb, 28(2), pp. 103-111.
- Bull, A. T., 2010. The renaissance of continuous culture in the post-genomics age. *J Ind Microbiol Biotechnol*, Oct, 37(10), pp. 993-1021.

- Bushell, M. E. et al., 2006. The use of genome scale metabolic flux variability analysis for process feed formulation based on an investigation of the effects of the zwf mutation on antibiotic production in *Streptomyces coelicolor*. *Enzyme and Microbial Technology*, 39(6), pp. 1347-1353.
- Butler, M. J. et al., 2002. Engineering of primary carbon metabolism for improved antibiotic production in *Streptomyces lividans*. *Appl Environ Microbiol*, Oct, 68(10), pp. 4731-4739.
- Chater, K. F. et al., 2010. The complex extracellular biology of *Streptomyces*. *FEMS Microbiol Rev*, Mar, 34(2), pp. 171-198.
- Chater, K. F. & Chandra, G., 2006. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev*, Sep, 30(5), pp. 651-672.
- Choi, S.-S., Kim, S.-H. & Kim, E.-S., 2010. Proteomics-driven identification of SCO4677-dependent proteins in *Streptomyces lividans* and *Streptomyces coelicolor*. *J Microbiol Biotechnol*, Mar, 20(3), pp. 480-484.
- Christensen, B., Gombert, A. K. & Nielsen, J., 2002. Analysis of flux estimates based on (13)C-labelling experiments. *Eur J Biochem*, Jun, 269(11), pp. 2795-2800.
- Christensen, B. & Nielsen, J., 2000. Metabolic network analysis of *Penicillium chrysogenum* using (13)C-labeled glucose. *Biotechnol Bioeng*, Jun, 68(6), pp. 652-659.
- Christensen, B. & Nielsen, J., 2000. Metabolic network analysis. A powerful tool in metabolic engineering. *Adv Biochem Eng Biotechnol*, Årgang 66, pp. 209-231.
- De Keersmaecker, S. C. J., Thijs, I. M. V., Vanderleyden, J. & Marchal, K., 2006. Integration of omics data: how well does it work for bacteria?. *Mol Microbiol*, Dec, 62(5), pp. 1239-1250.
- D'Huys, P.-J. et al., 2011. Amino acid uptake profiling of wild type and recombinant *Streptomyces lividans* TK24 batch fermentations. *J Biotechnol*, Apr, 152(4), pp. 132-143.
- D'Huys, P.-J. et al., 2012. Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. *J Biotechnol*, Sep, 161(1), pp. 1-13.
- Dietmair, S. et al., 2010. Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol. *Analytical Biochemistry*, 404(2), pp. 155-164.
- Dunn, W. B., 2008. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol*, 5(1), p. 011001.
- Durot, M., Bourguignon, P.-Y. & Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev*, Jan, 33(1), pp. 164-190.
- Dwarakanath, S. et al., 2012. Response to Copper Stress in *Streptomyces lividans* Extends beyond Genes under Direct Control of a Copper-sensitive Operon Repressor Protein (CsoR). *Journal of Biological Chemistry*, 287(21), pp. 17833-17847.

- Daae, E. B. & Ison, A. P., 1999. Classification and sensitivity analysis of a proposed primary metabolic reaction network for *Streptomyces lividans*.. *Metab Eng*, Apr, 1(2), pp. 153-165.
- Erxleben, A. et al., 2011. Genome sequence of *Streptomyces* sp. strain Tü6071.. *J Bacteriol*, Aug, 193(16), pp. 4278-4279.
- Escutia, M. R. et al., 2006. Compensatory effect of the minor *Streptomyces lividans* type I signal peptidases on the SipY major signal peptidase deficiency as determined by extracellular proteome analysis.. *Proteomics*, Jul, 6(14), pp. 4137-4146.
- Facey, P. D. et al., 2013. The Evolution of an Osmotically Inducible *dps* in the Genus *Streptomyces*.. *PLoS One*, 8(4), p. e60772.
- Feist, A. M. et al., 2009. Reconstruction of biochemical networks in microorganisms.. *Nat Rev Microbiol*, Feb, 7(2), pp. 129-143.
- Fiehn, O. et al., 2006. Establishing reporting standards for metabolomic and metabonomic studies: a call for participation.. *OMICS*, 10(2), pp. 158-163.
- Fischer, E., Zamboni, N. & Sauer, U., 2004. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ¹³C constraints.. *Anal Biochem*, Feb, 325(2), pp. 308-316.
- Flärdh, K. & Buttner, M. J., 2009. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium.. *Nat Rev Microbiol*, Jan, 7(1), pp. 36-49.
- Gernaey, K. V. et al., 2010. Application of mechanistic models to fermentation and biocatalysis for next-generation processes.. *Trends Biotechnol*, Jul, 28(7), pp. 346-354.
- Gombert, A. K., dos Santos, M. M., Christensen, B. & Nielsen, J., 2001. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression.. *J Bacteriol*, Feb, 183(4), pp. 1441-1451.
- Grubbs, K. J. et al., 2011. Genome sequence of *Streptomyces griseus* strain XylebKG-1, an ambrosia beetle-associated actinomycete.. *J Bacteriol*, Jun, 193(11), pp. 2890-2891.
- Gunnarsson, N., Mortensen, U. H., Sosio, M. & Nielsen, J., 2004. Identification of the Entner-Doudoroff pathway in an antibiotic-producing actinomycete species.. *Mol Microbiol*, May, 52(3), pp. 895-902.
- Hesketh, A. et al., 2007. New pleiotropic effects of eliminating a rare tRNA from *Streptomyces coelicolor*, revealed by combined proteomic and transcriptomic analysis of liquid cultures.. *BMC Genomics*, Årgang 8, p. 261.
- Hopwood, D. A., 1999. Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico.. *Microbiology*, Sep, Årgang 145 (Pt 9), pp. 2183-2202.
- Huang, J. et al., 2005. Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*.. *Mol Microbiol*, Dec, 58(5), pp. 1276-1287.

- Ikeda, H. et al., 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol*, May, 21(5), pp. 526-531.
- Jayapal, K. P. et al., 2007. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. *BMC Genomics*, Årgang 8, p. 229.
- Jonsbu, E., Christensen, B. & Nielsen, J., 2001. Changes of in vivo fluxes through central metabolic pathways during the production of nystatin by *Streptomyces noursei* in batch culture. *Appl Microbiol Biotechnol*, Jul, 56(1-2), pp. 93-100.
- Kassama, Y. et al., 2010. Assessment of adaptive focused acoustics versus manual vortex/freeze-thaw for intracellular metabolite extraction from *Streptomyces lividans* producing recombinant proteins using GC-MS and multi-block principal component analysis. *Analyst*, May, 135(5), pp. 934-942.
- Kauffman, K. J., Prakash, P. & Edwards, J. S., 2003. Advances in flux balance analysis. *Curr Opin Biotechnol*, Oct, 14(5), pp. 491-496.
- Kieser, H. M., Kieser, T. & Hopwood, D. A., 1992. A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. *J Bacteriol*, Sep, 174(17), pp. 5496-5507.
- Kim, C.-Y., Park, H.-J. & Kim, E.-S., 2005. Proteomics-driven identification of putative AfsR2-target proteins stimulating antibiotic biosynthesis in *Streptomyces lividans*. *Biotechnology and Bioprocess Engineering*, 10(3), pp. 248-253.
- King, R., 1997. A structured mathematical model for a class of organisms: I. Development of a model for *Streptomyces tendae* and application of model-based control. *Journal of Biotechnology*, 52(3), pp. 219-234.
- Kirk, S., Avignone-Rossa, C. A. & Bushell, M. E., 2000. Growth limiting substrate affects antibiotic production and associated metabolic fluxes in *Streptomyces clavuligerus*. *Biotechnology letters*, Årgang 22, p. 1803–1809.
- Kiviharju, K., Moilanen, U., Leisola, M. & Eerikäinen, T., 2007. A chemostat study of *Streptomyces peucetius* var. *caesius* N47. *Appl Microbiol Biotechnol*, Jan, 73(6), pp. 1267-1274.
- Kol, S. et al., 2010. Metabolomic characterization of the salt stress response in *Streptomyces coelicolor*. *Appl Environ Microbiol*, Apr, 76(8), pp. 2574-2581.
- Lee, J. M., Gianchandani, E. P. & Papin, J. A., 2006. Flux balance analysis in the era of metabolomics. *Brief Bioinform*, Jun, 7(2), pp. 140-150.
- Lewis, R. et al., 2010. Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. *BMC Genomics*, 11(1), p. 682.
- Li, H., Faury, D. & Morosoli, R., 2006. Impact of amino acid changes in the signal peptide on the secretion of the Tat-dependent xylanase C from *Streptomyces lividans*. *FEMS Microbiol Lett*, Feb, 255(2), pp. 268-274.

- Likic, V. A., McConville, M. J., Lithgow, T. & Bacic, A., 2010. Systems biology: the next frontier for bioinformatics.. *Adv Bioinformatics*, p. 268925.
- Lin, Y. B. et al., 2011. *Streptomyces zinciresistens* sp. nov., a zinc-resistant actinomycete isolated from soil from a copper and zinc mine.. *Int J Syst Evol Microbiol*, Mar, 61(Pt 3), pp. 616-620.
- Mahadevan, R. & Palsson, B. O., 2005. Properties of metabolic networks: structure versus function.. *Biophys J*, Jan, 88(1), pp. L07--L09.
- Maharjan, R. P. & Ferenci, T., 2003. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*.. *Anal Biochem*, Feb, 313(1), pp. 145-154.
- Manteca, A. et al., 2010. Quantitative proteomics analysis of *Streptomyces coelicolor* development demonstrates that onset of secondary metabolism coincides with hypha differentiation.. *Mol Cell Proteomics*, Jul, 9(7), pp. 1423-1436.
- Manteca, A. et al., 2008. Mycelium differentiation and antibiotic production in submerged cultures of *Streptomyces coelicolor*.. *Appl Environ Microbiol*, Jun, 74(12), pp. 3877-3886.
- Marioni, J. C. et al., 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.. *Genome Res*, Sep, 18(9), pp. 1509-1517.
- Martín, J. F. et al., 2012. Transcriptomic studies of phosphate control of primary and secondary metabolism in *Streptomyces coelicolor*.. *Appl Microbiol Biotechnol*, Jul, 95(1), pp. 61-75.
- McGuire, A. M. et al., 2012. Comparative analysis of *Mycobacterium* and related Actinomycetes yields insight into the evolution of *Mycobacterium tuberculosis* pathogenesis.. *BMC Genomics*, Årgang 13, p. 120.
- Nandy, S. K., Rattleff, S. & Lantz, A. E., 2013. Model of enhanced heterologous protein production by *Streptomyces lividans* – Effect of C/N ratio. *Manuscript in prep.*
- Nielsen, J., 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations.. *Biotechnol Bioeng*, 58(2-3), pp. 125-132.
- Nowruzi, K. et al., 2008. Development of a minimal defined medium for recombinant human interleukin-3 production by *Streptomyces lividans* 66. *Biotechnology and bioengineering*, 99(1), pp. 214-222.
- Nowruzi, K. et al., 2011. Metabolic flux-based optimisation of recombinant human interleukin-3 expression by *Streptomyces lividans* 66. *The Canadian Journal of Chemical Engineering*, 89(4), pp. 879-888.
- Oberhardt, M. A., Palsson, B. Ø. & Papin, J. A., 2009. Applications of genome-scale metabolic reconstructions.. *Mol Syst Biol*, Årgang 5, p. 320.
- Ohnishi, Y. et al., 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350.. *J Bacteriol*, Jun, 190(11), pp. 4050-4060.

- Omura, S. et al., 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites.. *Proc Natl Acad Sci U S A*, Oct, 98(21), pp. 12215-12220.
- Palsson, B., 2002. In silico biology through "omics".. *Nat Biotechnol*, Jul, 20(7), pp. 649-650.
- Paradkar, A., Trefzer, A., Chakraborty, R. & Stassi, D., 2003. *Streptomyces* genetics: a genomic perspective.. *Crit Rev Biotechnol*, 23(1), pp. 1-27.
- Price, N. D., Reed, J. L. & Palsson, B. Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints.. *Nat Rev Microbiol*, Nov, 2(11), pp. 886-897.
- Pullan, S. T., Chandra, G., Bibb, M. J. & Merrick, M., 2011. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes.. *BMC Genomics*, Årgang 12, p. 175.
- Rabinowitz, J. D. & Kimball, E., 2007. Acidic acetonitrile for cellular metabolome extraction from *Escherichia coli*.. *Anal Chem*, Aug, 79(16), pp. 6167-6173.
- Rattleff, S., Nielsen, J., Rokem, S. & Lantz, A. E., 2013b. Metabolic flux profiling of sixteen Actinobacteria species. *Manuscript in prep*.
- Reed, J. L., Famili, I., Thiele, I. & Palsson, B. O., 2006. Towards multidimensional genome annotation.. *Nat Rev Genet*, Feb, 7(2), pp. 130-141.
- Rossa, C. A. et al., 2002. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*.. *Metab Eng*, Apr, 4(2), pp. 138-150.
- Ryu, Y.-G., Butler, M. J., Chater, K. F. & Lee, K. J., 2006. Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*.. *Appl Environ Microbiol*, Nov, 72(11), pp. 7132-7139.
- Saito, N., Ohashi, Y., Soga, T. & Tomita, M., 2010. Unveiling cellular biochemical reactions via metabolomics-driven approaches.. *Curr Opin Microbiol*, Jun, 13(3), pp. 358-362.
- Sauer, U., 2006. Metabolic networks in motion: ¹³C-based flux analysis.. *Mol Syst Biol*, Årgang 2, p. 62.
- Schmidt, A. et al., 2009. Heavy metal resistance to the extreme: *Streptomyces* strains from a former uranium mining area. *Chemie der Erde - Geochemistry*, 69, Supplement 2(0), pp. 35-44.
- Sellick, C. A. et al., 2009. Effective quenching processes for physiologically valid metabolite profiling of suspension cultured Mammalian cells.. *Anal Chem*, Jan, 81(1), pp. 174-183.
- Studholme, D. J., Bentley, S. D. & Kormanec, J., 2004. Bioinformatic identification of novel regulatory DNA sequence motifs in *Streptomyces coelicolor*.. *BMC Microbiol*, Apr, Årgang 4, p. 14.

- van Veluw, G. J. et al., 2012. Analysis of two distinct mycelial populations in liquid-grown *Streptomyces* cultures using a flow cytometry-based proteomics approach.. *Appl Microbiol Biotechnol*, Dec, 96(5), pp. 1301-1312.
- van Wezel, G. P. et al., 2006. Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering.. *Appl Environ Microbiol*, Aug, 72(8), pp. 5283-5288.
- Varma, A. & Palsson, B. O., 1994. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nature Biotechnology*, Årgang 12, pp. 994-998.
- Ventura, M. et al., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum.. *Microbiol Mol Biol Rev*, Sep, 71(3), pp. 495-548.
- Villas-Bôas, S. G. & Bruheim, P., 2007. Cold glycerol-saline: the promising quenching solution for accurate intracellular metabolite analysis of microbial cells.. *Anal Biochem*, Nov, 370(1), pp. 87-97.
- Villas-Bôas, S. G. et al., 2005. Global metabolite analysis of yeast: evaluation of sample preparation methods.. *Yeast*, Oct, 22(14), pp. 1155-1169.
- Villas-Bôas, S. G. et al., 2006. Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media.. *Anal Biochem*, Feb, 349(2), pp. 297-305.
- Wang, X.-J. et al., 2010. Genome sequence of the milbemycin-producing bacterium *Streptomyces bingchengensis*.. *J Bacteriol*, Sep, 192(17), pp. 4526-4527.
- Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics.. *Nat Rev Genet*, Jan, 10(1), pp. 57-63.
- Wittmann, C. et al., 2004. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria.. *Anal Biochem*, Apr, 327(1), pp. 135-139.
- Yagüe, P. et al., 2013. Transcriptomic analysis of *Streptomyces coelicolor* differentiation in solid sporulating cultures: first compartmentalized and second multinucleated mycelia have different and distinctive transcriptomes.. *PLoS One*, 8(3), p. e60665.
- Yaxley, A. M., 2009. *Study of the complete genome sequence of Streptomyces scabies (or scabiei) 87.22*, s.l.: s.n.
- Zhang, W., Li, F. & Nie, L., 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies.. *Microbiology*, Feb, 156(Pt 2), pp. 287-301.
- Zhou, Z., Gu, J., Li, Y.-Q. & Wang, Y., 2012. Genome plasticity and systems evolution in *Streptomyces*.. *BMC Bioinformatics*, Årgang 13 Suppl 10, p. S8.

Chapter 7. Discussion and perspectives

While much research has gone into implementing new hosts for heterologous protein production, and *S. lividans* has taken a promising place among the *Actinomycete* hosts, work is still required before it can be considered a generally applicable host. This thesis is the result of a project that has been connected to an EU project named STREPTOMICS, with the specific purpose of introducing *S. lividans* among the general hosts for heterologous protein expression. Many of the prerequisites mentioned in chapter 1 are already in place, products produced in *S. lividans* are available, and despite the few actual products, it does indicate that *S. lividans* is capable of use in an industrial setting. Some of the work that needs to be performed is gaining a better understanding of the quantitative physiology of strains of *S. lividans* producing various industrially relevant proteins.

Summing up

The basic quantitative physiology has been studied in the chapters 3 and 4 of this thesis. Further an approach that is compatible with high throughput methods, has been used to screen a large number of carbon and nitrogen sources for their influence on growth and product formation. The effect of the carbon sources seemed polar, either a high amount of growth or a high amount of product, whereas with the nitrogen sources a number of n-sources resulted in high amount of growth and productivity. These were tested in larger scales, where more quantitative data could be obtained at the expense of the amount of studied sources. The study of the nitrogen sources resulted in a paper submitted to Applied Microbiology and Biotechnology. The ultimately best sources were aspartate and glutamate, even beyond normalising for carbon contribution. Ammonium seems to be a mediocre nitrogen source, on par with nitrate, though when using nitrate more by-products were found, thus making ammonium more preferable.

The quantitative physiology was more thoroughly studied in chapter 5, where chemostats were used to study the effect of growth rate on product formation. Subjecting *S. lividans* to extremely low dilution rates, resulted in a forced production of actinorhodin, but only at dilution rates of less than $D \leq 0.035 \text{ h}^{-1}$. Morphologically it has been observed that at medium to low dilution rates the WT would grow freely dispersed, whereas in the batch experiments, the WT would always form pellets. The mRFP strain for unknown reasons obtained higher growth rates in batch, and displayed a freely dispersed morphology at all growth/dilution rates. It was seen that protein production was coupled to growth rate, i.e. the higher the growth rate the higher the productivity. Within the range of growth rates studied no upper limitation was found.

As supporting foundation of chapter 6 are two papers that have yet to be submitted. One is the preliminary flux analysis of several actinomycetes to assess ED-pathway activity, using labelled substrates. The other is a co-authorship with Subir Kumar Nandy, of the development of an unstructured kinetic model to predict growth and product formation taking both carbon and nitrogen source into account. This model has both been used with two and three substrates (one carbon and one nitrogen source, or one carbon-, one nitrogen-, and one mixed source [i.e. an amino acid]), see chapter 6. The purpose was to estimate the effect of adjusting the C:N ratio, also when comparing to mixed sources as amino acids. The findings

suggest that a higher productivity can be obtained by addition of more nitrogen, and from the results the better productivity was observed for ammonium relative to ammonium glutamate mix. In order to optimize comparability a fixed amount of glutamate was chosen (to maintain a fixed total level of carbon), and the C:N ratio was adjusted by the addition of NH_4Cl .

The physiology of *S. lividans* has been described across several laboratory relevant volumes of fermentations. A degree of variability has been observed that suggests more research in strain propagation is necessary, as well as it would be prudent to test the expression from chromosomally integrated genes.

Perspectives

The fields of research that are currently being expanded regarding *S. lividans*, and related strains, include development of genetic tools, the study of morphology and developmental biology, fermentation knowledge across scales (including this thesis), and genomics with availability of 18 completely and 26 partially sequenced and/or annotated genome sequences within the genus *Streptomyces*.

Genomics

Genomics open for some interesting fields of study, some of which are already being examined. With the publication of genome sequences for *S. coelicolor* (Bentley, et al. 2002) and *S. avermitilis* (Ikeda, et al. 2003) came also the realisation that a large number of so-called cryptic secondary metabolites were encoded. This enabled genome mining for unknown metabolites, though it was realised that it was necessary to move these clusters to a strain with a clean background. The relevant developments here includes, artificial chromosomes for moving large segments of DNA (Alduina og Gallo 2012), as well as the various approaches used to produce strains with clean backgrounds, i.e. strains which produce little or no endogenous secondary metabolites. The ‘cleaning’ of these strains was performed in one of three ways: I) sequentially deleting portions of the arms of the chromosomes, which are recognised to carry the largest part of the secondary metabolites. II) identification and sequential deletion of secondary metabolite synthases (NRPS synthases and polyketide synthases), an approach that has proved useful in *S. coelicolor* (Gomez-Escribano og Bibb 2011). III) the study and use of regulation. *Streptomyces* have an extraordinary amount of (putative) regulatory proteins, and mapping the regulatory network, would allow deleting/overexpressing a set of regulators, and potentially obtaining a strain that produces no secondary metabolites, with fewer changes than used by method II. Strains with their background “cleaned” would likewise be promising to use as producers of primary metabolites, since little carbon would be lost to secondary metabolism by-products, and purification/DSP would be eased; further a clean background would also be considered safer, in a governmental regulatory view, if producing human/pharmaceutical products for human use.

Regulation

Genomics of regulatory processes are an entrance to investigate, and possibly solve, several challenges related to *Actinomycetes* including, as mentioned above, reducing by-products from secondary metabolism. Describing and characterising the regulatory network would possibly help control the strains under fermentation conditions to benefit both primary and secondary metabolite producers. The ability to model the behaviour of the cell in changing conditions would improve the descriptive power of the model, as mentioned in chapter 6. (Baltz 2011) reviews some of these methods, and shows several examples where heterologous overexpression of a positive regulator, or disrupting a negative regulator, can result in high production of secondary metabolites. In the case of primary metabolism products, the opposite

method could be used as an alternative to, or in extension of, the sequential deletion method to produce a strain sporting a cleaner background, and less competing products. Potentially regulation could be used to understand and reduce, or benefit from, stress responses, e.g. unfolded protein response. Otherwise could regulation of amino acid catabolism or protein degradation be targeted for improving protein yield. Furthermore (van Wezel, et al. 2006) has in a morphological engineering approach tested several strains of *Streptomyces* overexpressing the morphogene *ssgA*, which results in a high degree of branching, and higher enzyme production. In a later paper (Noens, et al. 2007), describes this gene as a cell cycle control protein, and states that overexpression of the protein can result in submerged sporulation. Submerged sporulation is unwanted, but the increased branching and protein productivity obtained from overexpressing *ssgA* outlines some of the promise presented by better understanding the regulatory network.

Modelling and 'omics'

Modelling is a field that shows great promise for obtaining a deeper understanding of the relationship between an observed phenotype, and the genetic foundation that causes this. Integrative models are emerging where genome scale models are being coupled to time-series measurements of other 'omics' data; one such model has been used to describe the transcriptomic changes of *S. coelicolor* over the metabolic switch from primary to secondary metabolism (Alam, et al. 2010). A different kind of model has been developed to understand the influence of morphology on production (Celler, et al. 2012). These models are likewise at present unconnected to the genetic foundation, though obtaining this connection would result in strong models, with improved descriptive power.

The fact that two point mutations *rpoB* and *rpsL* can result in higher production of secondary metabolites (Jayapal, et al. 2010) also shows that there is a lot we still do not know, and are not really able to model yet. The reasons why these two point mutations increase secondary metabolism is not known, but they affect the ribosome, which causes more protein production in the secondary metabolism. The full extent of the pleiotropic effects of these point mutations have so far not been fully described, though they would be hard to model, since the molecular knowledge of the ribosome, however much it has been studied, is not completely understood. Furthermore it would result in extremely computer intensive models, if all elements in a genome scale network would have to be modelled at the molecular level. These two point mutations represent an example where it is high to impossible to foresee the effect and illustrates that methods known from classical strain optimisation may still be of excellent use and are complementary to the more modern, genetic engineering methods.

Classical vs. modern methods

Historically random mutagenesis has been used to improve production of either primary products like amino acids and organic acids, as well as products from the secondary metabolism. One example of sequential mutagenesis has increased antibiotic production in a strain of *S. albus* from 0.25 g/L to 10 g/L, and further to 15 g/L with a directed application of the *rpsL* mutation. This clearly suggests that some of the classical strain optimisation routines are not yet obsolete, and some of the methods have been refined to produce e.g. laboratory scale evolution (directed evolution) (Charusanti 2012).

Directed evolution is a rational approach which aims at obtaining the beneficial mutations (randomly) to cope with an increasing selective pressure, and afterwards the resulting strain can be sequenced, either full genome or selected regions of the genome, to try to understand what happened. Potentially knowledge

from such experiments will lead to increased molecular understanding, thus benefitting both basic research as well as the specific product, should the strain become publicly available. The difficult part in directed evolution is choosing the proper selective criteria, which will allow the strain to improve production of the wanted product. This can be achieved for strong selective pressures, like resistance to environmentally harsh conditions, e.g. extreme temperatures/pH or solvent/salt concentrations. With proteins that can potentially release nutrient from otherwise inaccessible sources, e.g. xyloglucanase/cellulases etc., selection can be introduced as limitations in nutrients, only provided by the specific source, though cellular optimisation may be improvement of transporters, or changes in the protein itself, which may be undesired. It is a method that shows great promise, especially for the secondary metabolite products, though the method has been used for optimising primary metabolism products. Directed evolution is a useful extension of genetic engineering, since it allows recovery of the growth rate after genetically modifying the cell.

Genetic stability

Genetic stability was studied in the 80's, and the plasmid stability of the pIJ101 plasmid was assessed. However, it is still an issue, especially since one of the obvious optimisation criteria for heterologous expression is integrating several copies of the gene of interest, followed by using strong promoters. The rationale is that this will ensure a higher level of available transcript, which potentially can result in higher product formation, though the relation between transcript and protein is not linear. Several copies of the genes increase the genetic instability due to homologous recombination, resulting in a strain that is unviable for industrial use (Baltz 2011). One method to circumvent gene loss is utilising integration sites flanked by essential genes. This way loop-out can be effectively eliminated, as in the case of (Mikkelsen, et al. 2012). This method would of course suggest, like for the genome minimised strain or the cleaned strain, creating a library of strains prepared for industrial use, with selected and characterised integration sites already prepared. The method also highlights the importance of knowledge transfer from entirely separate domains, such as yeast and filamentous fungi. Filamentous fungi have some of the same features/challenges as *Streptomyces* with respect to morphology, while yeast is very well studied and methods are developed and tested in a fast pace; therefore the most promising technologies can be transferred to *Streptomyces*, such as the integration strategy by (Mikkelsen, et al. 2012) mentioned above.

While a large number of vectors exist for heterologous expression in *Streptomyces*, many of them are multicopy plasmids. One problem with this approach is that multicopy plasmids introduce an extra level of population heterogeneity; there is already a great level of heterogeneity, since cells are found in several growth stages, with each their specific penchant for producing enzymes/secondary metabolites/spores, especially when large clumps or pellets are present. Adding variability in the number of plasmid per cell, the multicopy plasmids that are pIJ101 derived have on average between 40 and 100 copies, unnecessarily complicates the analyses of the data. Ensuring an equal level of transcript across the population would greatly reduce complications, and risk of plasmid loss can also be avoided by using integrative plasmids when establishing a production strain; which is already a known fact, though for cursory tests plasmid expression can suffice. Phage integration sites, and vectors utilising these, are available and could be used for chromosomal integration of heterologous genes, to increase strain genetic stability (Baltz 2011). Some of the phage sites are located in the stable core region of the chromosome (e.g. ϕ C31 or ϕ BT1) (Baltz 2011)

Fermentation conditions

Further (Alduina og Gallo 2012) outlines the importance of optimising the fermentation conditions, e.g. by feeding precursors for the products, or by changing the concentration of the limiting nutrient. Alternatively the choice, and ratio, of C- and N-sources greatly affect protein production and cellular regulation, and are therefore important to optimise likewise. These are important steps for the production of primary metabolites. As a metabolic engineering strategy for improving production, the first step may well be to feed precursors, followed by the second step, which would be to genetically alter the cell to improve precursor supply. With heterologous products spanning both bulk and high value-added products spheres; financially some products simply do not allow addition of defined components to the medium. Fermentation optimisation does also include optimising the physical parameters, in a more traditional DoE approach, adjusting e.g. pH, temperature, stirring, aeration and nutrient supply. I have introduced one method, which is compatible with high throughput automation, which allows screening large numbers of carbon/nitrogen/phosphorous/sulphur sources for effect on growth and product formation. This method is presented in chapters 3 and 4, where the latter has been submitted as a paper to AMB. Further, to fully be able to exploit the 'omics' methods, the reproducibility of the experiments have to be high. This can be obtained either by thoroughly defined batch experiments in bioreactors; this has been established in the chapters 3 and 4, or by the application of continuous cultivations, e.g. chemostats, where highly controlled highly reproducible samples can be obtained, for improving the reliability of 'omics' analyses.

Model strains

Streptomyces lividans, as well as *S. coelicolor* have become the model organisms for heterologous expression in the *Streptomyces* genus, as for several other *Actinomycetes*. *S. lividans* is known for its relaxed restriction system, low antibiotic production and remarkably low protease activity, whereas *S. coelicolor* has become the model for antibiotic production, and developmental biology, as well as host for several heterologous antibiotic clusters. In the heterologous expression of secondary metabolites, some methods have been developed that could be beneficial for primary products as well. Artificial chromosomes, as reviewed by (Alduina og Gallo 2012), have promised to ease the movement of large DNA fragments, especially with regard to secondary metabolite clusters. In the same paper they describe several methods for producing clean background strains, as has successfully been applied for *S. coelicolor*.

Physiological studies are among the studies that are still needed in developing *S. lividans* as a production host. Recently this avenue of research has been followed by more groups, and advanced analyses, like flux (D'Huys, et al. 2011) and time-course transcriptomics (Alam, et al. 2010) are appearing with greater frequency. (Anné, et al. 2012) published a review, as a summation of what was achieved in the STREPTOMICS project, which also contained the outlook on further developments needed. They likewise point out that physiological characterisation, as well as characterisation of the transport machinery and signal peptides are of key importance in the final developing steps.

Concluding remarks.

The work in this thesis has primarily focused on developing the methods to be used for obtaining the best data from the advanced 'omics' analyses. To achieve this, a thorough physiology platform had to be established. Firstly basic quantitative physiology was to be estimated, and reproducible conditions in a batch reactor had to be obtained. This was supported by a screening set up, where the optimum nutrient sources were identified. Subsequently this knowledge was applied to design continuous cultivations, that

could reliably and reproducibly be sampled for advanced 'omics' analyses. Further, the use of chemostats entirely eliminated the effect of the growth rate when comparing mutants, under comparable conditions. Thereby the optimal benefit can be gained in the future from the advanced analyses, since the data will be robust, the conditions controlled, the experiments repeatable. Lastly a deeper understanding of the organism was, and in future experiment will be, obtained, and possibly underlying regulatory events can be uncovered.

Chapter 8. Appendix A –

Additional carbon source figures

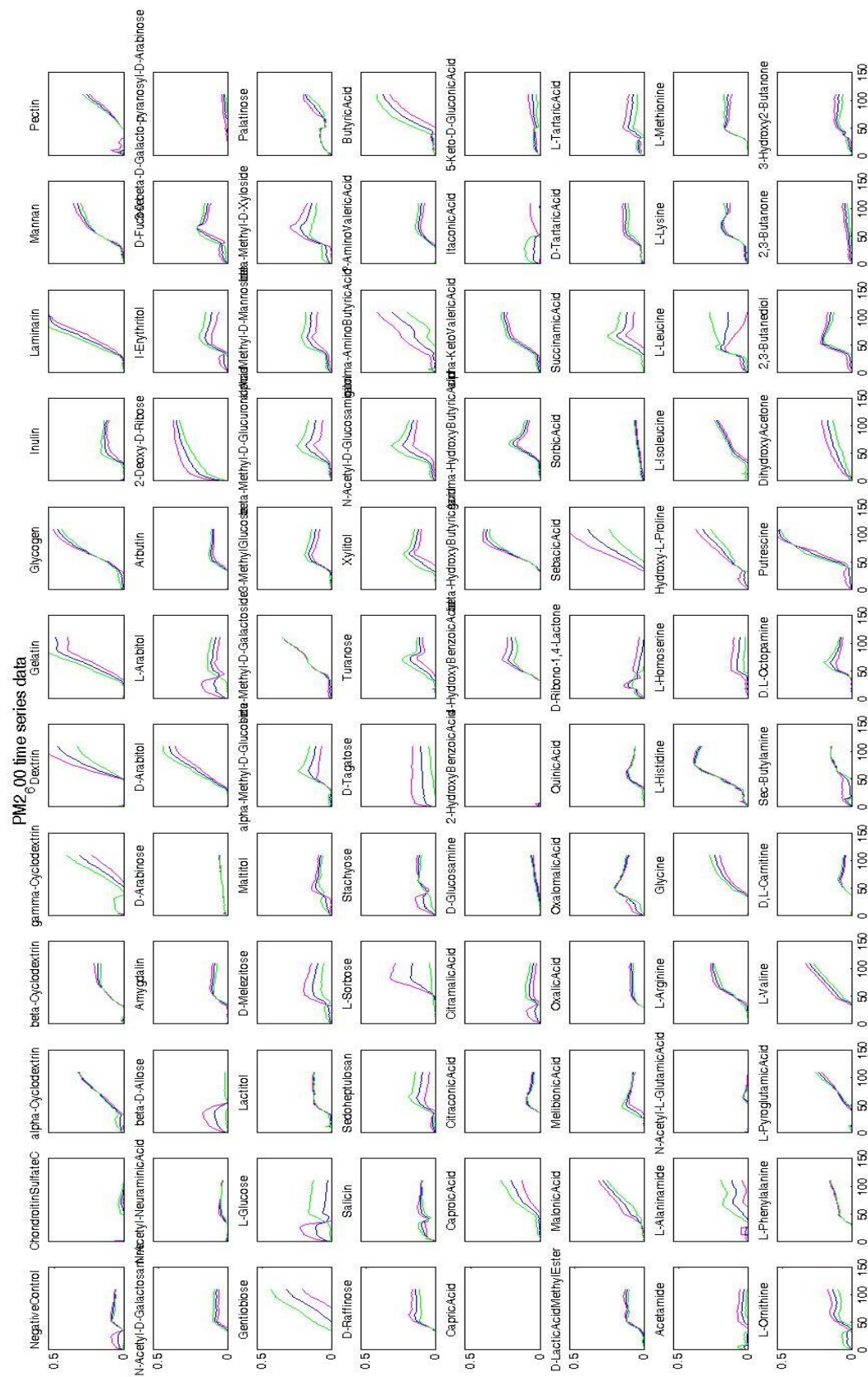


Figure 24 PM02 time course OD vs time. The magenta and green lines represent the individual plates, and the blue line is the average between these two plates, the negative control has been marked with a red ring, and with green rings are glucose and glutamate

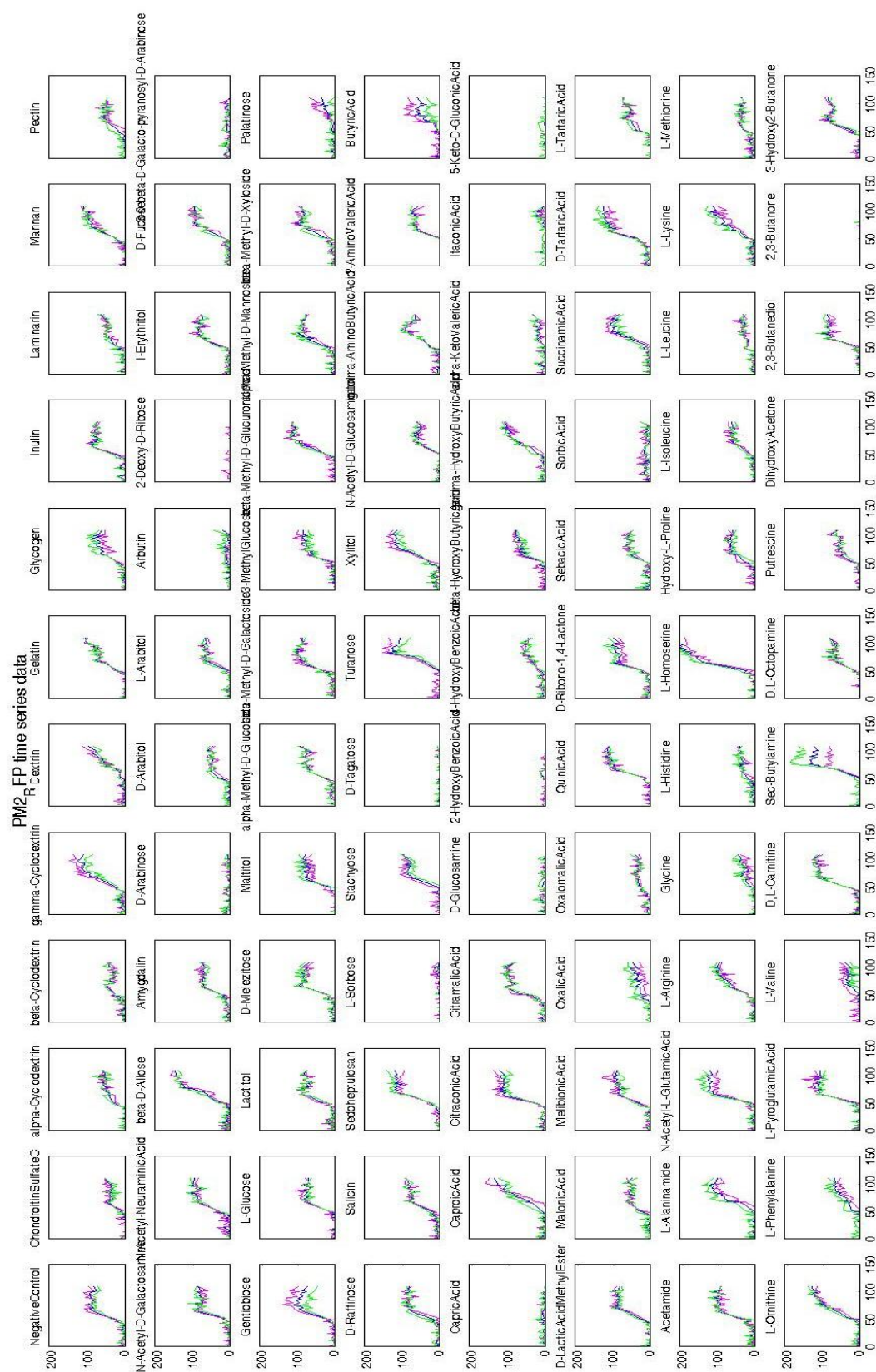


Figure 25 PM02 time course RFP vs time. The magenta and green lines represent the individual plates, and the blue line is the average between these two plates, the negative control has been marked with a red ring, and with green rings are glucose and glutamate

Chapter 9. Appendix B – Table of contents of Biolog plates

Figure 26 The list of substrates in the plates PM01, PM02, PM03 and PM04, referred from a text file received from Biolog

PM01	A01	Negative Control	C-Source
PM01	A02	L-Arabinose	C-Source
PM01	A03	N-Acetyl-D-Glucosamine	C-Source
PM01	A04	D-Saccharic Acid	C-Source
PM01	A05	Succinic Acid	C-Source
PM01	A06	D-Galactose	C-Source
PM01	A07	L-Aspartic Acid	C-Source
PM01	A08	L-Proline	C-Source
PM01	A09	D-Alanine	C-Source
PM01	A10	D-Trehalose	C-Source
PM01	A11	D-Mannose	C-Source
PM01	A12	Dulcitol	C-Source
PM01	B01	D-Serine	C-Source
PM01	B02	D-Sorbitol	C-Source
PM01	B03	Glycerol	C-Source
PM01	B04	L-Fucose	C-Source
PM01	B05	D-Glucuronic Acid	C-Source
PM01	B06	D-Gluconic Acid	C-Source
PM01	B07	D,L- α -Glycerol Phosphate	C-Source
PM01	B08	D-Xylose	C-Source
PM01	B09	D,L-Lactic acid	C-Source
PM01	B10	Formic Acid	C-Source
PM01	B11	D-Mannitol	C-Source
PM01	B12	L-Glutamic Acid	C-Source
PM01	C01	D-Glucose-6-Phosphate	C-Source
PM01	C02	D-Galactonic Acid-g-Lactone	C-Source
PM01	C03	D,L-Malic Acid	C-Source
PM01	C04	D-Ribose	C-Source
PM01	C05	Tween 20	C-Source
PM01	C06	L-Rhamnose	C-Source
PM01	C07	D-Fructose	C-Source
PM01	C08	Acetic Acid	C-Source
PM01	C09	α -D-Glucose	C-Source
PM01	C10	Maltose	C-Source
PM01	C11	D-Melibiose	C-Source
PM01	C12	Thymidine	C-Source
PM01	D01	L-Asparagine	C-Source
PM01	D02	D-Aspartic Acid	C-Source
PM01	D03	D-Glucosaminic Acid	C-Source
PM01	D04	1,2-Propanediol	C-Source
PM01	D05	Tween 40	C-Source
PM01	D06	α -Ketoglutaric Acid	C-Source

PM01	D07	a-Ketobutyric Acid	C-Source
PM01	D08	a-Methyl-D-Galactoside	C-Source
PM01	D09	a-D-Lactose	C-Source
PM01	D10	Lactulose	C-Source
PM01	D11	Sucrose	C-Source
PM01	D12	Uridine	C-Source
PM01	E01	L-Glutamine	C-Source
PM01	E02	m-Tartaric Acid	C-Source
PM01	E03	D-Glucose-1-Phosphate	C-Source
PM01	E04	D-Fructose-6-Phosphate	C-Source
PM01	E05	Tween 80	C-Source
PM01	E06	a-Hydroxyglutaric Acid-g-Lactone	C-Source
PM01	E07	a-Hydroxybutyric Acid	C-Source
PM01	E08	b-Methyl-D-Glucoside	C-Source
PM01	E09	Adonitol	C-Source
PM01	E10	Maltotriose	C-Source
PM01	E11	2'-Deoxyadenosine	C-Source
PM01	E12	Adenosine	C-Source
PM01	F01	Gly-Asp	C-Source
PM01	F02	Citric Acid	C-Source
PM01	F03	m-Inositol	C-Source
PM01	F04	D-Threonine	C-Source
PM01	F05	Fumaric Acid	C-Source
PM01	F06	Bromosuccinic Acid	C-Source
PM01	F07	Propionic Acid	C-Source
PM01	F08	Mucic Acid	C-Source
PM01	F09	Glycolic Acid	C-Source
PM01	F10	Glyoxylic Acid	C-Source
PM01	F11	D-Cellobiose	C-Source
PM01	F12	Inosine	C-Source
PM01	G01	Gly-Glu	C-Source
PM01	G02	Tricarballic Acid	C-Source
PM01	G03	L-Serine	C-Source
PM01	G04	L-Threonine	C-Source
PM01	G05	L-Alanine	C-Source
PM01	G06	Ala-Gly	C-Source
PM01	G07	Acetoacetic Acid	C-Source
PM01	G08	N-Acetyl-D-Mannosamine	C-Source
PM01	G09	Mono-Methylsuccinate	C-Source
PM01	G10	Methylpyruvate	C-Source
PM01	G11	D-Malic Acid	C-Source
PM01	G12	L-Malic Acid	C-Source
PM01	H01	Gly-Pro	C-Source
PM01	H02	p-Hydroxyphenyl Acetic Acid	C-Source
PM01	H03	m-Hydroxyphenyl Acetic Acid	C-Source
PM01	H04	Tyramine	C-Source
PM01	H05	D-Psicose	C-Source
PM01	H06	L-Lyxose	C-Source
PM01	H07	Glucuronamide	C-Source
PM01	H08	Pyruvic Acid	C-Source
PM01	H09	L-Galactonic Acid-g-Lactone	C-Source
PM01	H10	D-Galacturonic Acid	C-Source

PM01	H11	b-Phenylethylamine	C-Source
PM01	H12	2-Aminoethanol	C-Source
PM02A	A01	Negative Control	C-Source
PM02A	A02	Chondroitin Sulfate C	C-Source
PM02A	A03	a-Cyclodextrin	C-Source
PM02A	A04	b-Cyclodextrin	C-Source
PM02A	A05	g-Cyclodextrin	C-Source
PM02A	A06	Dextrin	C-Source
PM02A	A07	Gelatin	C-Source
PM02A	A08	Glycogen	C-Source
PM02A	A09	Inulin	C-Source
PM02A	A10	Laminarin	C-Source
PM02A	A11	Mannan	C-Source
PM02A	A12	Pectin	C-Source
PM02A	B01	N-Acetyl-D-Galactosamine	C-Source
PM02A	B02	N-Acetyl-Neuraminic Acid	C-Source
PM02A	B03	b-D-Allose	C-Source
PM02A	B04	Amygdalin	C-Source
PM02A	B05	D-Arabinose	C-Source
PM02A	B06	D-Arabitol	C-Source
PM02A	B07	L-Arabitol	C-Source
PM02A	B08	Arbutin	C-Source
PM02A	B09	2-Deoxy-D-Ribose	C-Source
PM02A	B10	i-Erythritol	C-Source
PM02A	B11	D-Fucose	C-Source
PM02A	B12	3-O-?-D-Galactopyranosyl-D-Arabinose	C-Source
PM02A	C01	Gentiobiose	C-Source
PM02A	C02	L-Glucose	C-Source
PM02A	C03	D-Lactitol	C-Source
PM02A	C04	D-Melezitose	C-Source
PM02A	C05	Maltitol	C-Source
PM02A	C06	a-Methyl-D-Glucoside	C-Source
PM02A	C07	b-Methyl-D-Galactoside	C-Source
PM02A	C08	3-Methylglucose	C-Source
PM02A	C09	b-Methyl-D-Glucuronic Acid	C-Source
PM02A	C10	a-Methyl-D-Mannoside	C-Source
PM02A	C11	b-Methyl-D-Xyloside	C-Source
PM02A	C12	Palatinose	C-Source
PM02A	D01	D-Raffinose	C-Source
PM02A	D02	Salicin	C-Source
PM02A	D03	Sedoheptulosan	C-Source
PM02A	D04	L-Sorbose	C-Source
PM02A	D05	Stachyose	C-Source
PM02A	D06	D-Tagatose	C-Source
PM02A	D07	Turanose	C-Source
PM02A	D08	Xylitol	C-Source
PM02A	D09	N-Acetyl-D-glucosaminitol	C-Source
PM02A	D10	g-Amino-N-Butyric Acid	C-Source
PM02A	D11	d-Amino Valeric Acid	C-Source
PM02A	D12	Sodium butyrate	C-Source
PM02A	E01	Capric Acid	C-Source
PM02A	E02	Caproic Acid	C-Source

PM02A	E03	Citraconic Acid	C-Source
PM02A	E04	D,L-Citramalic Acid	C-Source
PM02A	E05	D-Glucosamine	C-Source
PM02A	E06	2-Hydroxybenzoic acid	C-Source
PM02A	E07	4-Hydroxybenzoic Acid	C-Source
PM02A	E08	b-Hydroxybutyric Acid	C-Source
PM02A	E09	g-Hydroxybutyric Acid	C-Source
PM02A	E10	2-Oxovaleric acid	C-Source
PM02A	E11	Itaconic Acid	C-Source
PM02A	E12	5-Keto-D-Gluconic Acid	C-Source
PM02A	F01	D-Lactic Acid Methyl Ester	C-Source
PM02A	F02	Malonic Acid	C-Source
PM02A	F03	Melibionc Acid	C-Source
PM02A	F04	Oxalic Acid	C-Source
PM02A	F05	Oxalomalic Acid	C-Source
PM02A	F06	Quinic Acid	C-Source
PM02A	F07	D-Ribono-1,4-Lactone	C-Source
PM02A	F08	Sebacic Acid	C-Source
PM02A	F09	Sorbic Acid	C-Source
PM02A	F10	Succinamic Acid	C-Source
PM02A	F11	D-Tartaric Acid	C-Source
PM02A	F12	L-Tartaric Acid	C-Source
PM02A	G01	Acetamide	C-Source
PM02A	G02	L-Alaninamide	C-Source
PM02A	G03	N-Acetyl-L-Glutamic Acid	C-Source
PM02A	G04	L-Arginine	C-Source
PM02A	G05	Glycine	C-Source
PM02A	G06	L-Histidine	C-Source
PM02A	G07	L-Homoserine	C-Source
PM02A	G08	Hydroxy-L-Proline	C-Source
PM02A	G09	L-Isoleucine	C-Source
PM02A	G10	L-Leucine	C-Source
PM02A	G11	L-Lysine	C-Source
PM02A	G12	L-Methionine	C-Source
PM02A	H01	L-Ornithine	C-Source
PM02A	H02	L-Phenylalanine	C-Source
PM02A	H03	L-Pyroglutamic Acid	C-Source
PM02A	H04	L-Valine	C-Source
PM02A	H05	D,L-Carnitine	C-Source
PM02A	H06	Sec-Butylamine	C-Source
PM02A	H07	D,L-Octopamine	C-Source
PM02A	H08	Putrescine	C-Source
PM02A	H09	Dihydroxyacetone	C-Source
PM02A	H10	2,3-Butanediol	C-Source
PM02A	H11	2,3-Butanone	C-Source
PM02A	H12	3-Hydroxy 2-Butanone	C-Source
PM03B	A01	Negative Control	N-Source
PM03B	A02	Ammonia	N-Source
PM03B	A03	Nitrite	N-Source
PM03B	A04	Nitrate	N-Source
PM03B	A05	Urea	N-Source
PM03B	A06	Biuret	N-Source

PM03B	A07	L-Alanine	N-Source
PM03B	A08	L-Arginine	N-Source
PM03B	A09	L-Asparagine	N-Source
PM03B	A10	L-Aspartic Acid	N-Source
PM03B	A11	L-Cysteine	N-Source
PM03B	A12	L-Glutamic Acid	N-Source
PM03B	B01	L-Glutamine	N-Source
PM03B	B02	Glycine	N-Source
PM03B	B03	L-Histidine	N-Source
PM03B	B04	L-Isoleucine	N-Source
PM03B	B05	L-Leucine	N-Source
PM03B	B06	L-Lysine	N-Source
PM03B	B07	L-Methionine	N-Source
PM03B	B08	L-Phenylalanine	N-Source
PM03B	B09	L-Proline	N-Source
PM03B	B10	L-Serine	N-Source
PM03B	B11	L-Threonine	N-Source
PM03B	B12	L-Tryptophan	N-Source
PM03B	C01	L-Tyrosine	N-Source
PM03B	C02	L-Valine	N-Source
PM03B	C03	D-Alanine	N-Source
PM03B	C04	D-Asparagine	N-Source
PM03B	C05	D-Aspartic Acid	N-Source
PM03B	C06	D-Glutamic Acid	N-Source
PM03B	C07	D-Lysine	N-Source
PM03B	C08	D-Serine	N-Source
PM03B	C09	D-Valine	N-Source
PM03B	C10	L-Citrulline	N-Source
PM03B	C11	L-Homoserine	N-Source
PM03B	C12	L-Ornithine	N-Source
PM03B	D01	N-Acetyl-L-Glutamic Acid	N-Source
PM03B	D02	N-Phthaloyl-L-Glutamic Acid	N-Source
PM03B	D03	L-Pyroglutamic Acid	N-Source
PM03B	D04	Hydroxylamine	N-Source
PM03B	D05	Methylamine	N-Source
PM03B	D06	N-Amylamine	N-Source
PM03B	D07	N-Butylamine	N-Source
PM03B	D08	Ethylamine	N-Source
PM03B	D09	Ethanolamine	N-Source
PM03B	D10	Ethylenediamine	N-Source
PM03B	D11	Putrescine	N-Source
PM03B	D12	Agmatine	N-Source
PM03B	E01	Histamine	N-Source
PM03B	E02	b-Phenylethylamine	N-Source
PM03B	E03	Tyramine	N-Source
PM03B	E04	Acetamide	N-Source
PM03B	E05	Formamide	N-Source
PM03B	E06	Glucuronamide	N-Source
PM03B	E07	D,L-Lactamide	N-Source
PM03B	E08	D-Glucosamine	N-Source
PM03B	E09	D-Galactosamine	N-Source
PM03B	E10	D-Mannosamine	N-Source

PM03B	E11	N-Acetyl-D-Glucosamine	N-Source
PM03B	E12	N-Acetyl-D-Galactosamine	N-Source
PM03B	F01	N-Acetyl-D-Mannosamine	N-Source
PM03B	F02	Adenine	N-Source
PM03B	F03	Adenosine	N-Source
PM03B	F04	Cytidine	N-Source
PM03B	F05	Cytosine	N-Source
PM03B	F06	Guanine	N-Source
PM03B	F07	Guanosine	N-Source
PM03B	F08	Thymine	N-Source
PM03B	F09	Thymidine	N-Source
PM03B	F10	Uracil	N-Source
PM03B	F11	Uridine	N-Source
PM03B	F12	Inosine	N-Source
PM03B	G01	Xanthine	N-Source
PM03B	G02	Xanthosine	N-Source
PM03B	G03	Uric Acid	N-Source
PM03B	G04	Alloxan	N-Source
PM03B	G05	Allantoin	N-Source
PM03B	G06	Parabanic Acid	N-Source
PM03B	G07	D,L-α-Amino-N-Butyric Acid	N-Source
PM03B	G08	γ-Amino-N-Butyric Acid	N-Source
PM03B	G09	ε-Amino-N-Caproic Acid	N-Source
PM03B	G10	D,L-α-Amino-Caprylic Acid	N-Source
PM03B	G11	d-Amino-N-Valeric Acid	N-Source
PM03B	G12	α-Amino-N-Valeric Acid	N-Source
PM03B	H01	Ala-Asp	N-Source
PM03B	H02	Ala-Gln	N-Source
PM03B	H03	Ala-Glu	N-Source
PM03B	H04	Ala-Gly	N-Source
PM03B	H05	Ala-His	N-Source
PM03B	H06	Ala-Leu	N-Source
PM03B	H07	Ala-Thr	N-Source
PM03B	H08	Gly-Asn	N-Source
PM03B	H09	Gly-Gln	N-Source
PM03B	H10	Gly-Glu	N-Source
PM03B	H11	Gly-Met	N-Source
PM03B	H12	Met-Ala	N-Source
PM04A	A01	Negative Control	P-Source
PM04A	A02	Phosphate	P-Source
PM04A	A03	Pyrophosphate	P-Source
PM04A	A04	Trimetaphosphate	P-Source
PM04A	A05	Tripolyphosphate	P-Source
PM04A	A06	Triethyl Phosphate	P-Source
PM04A	A07	Hypophosphite	P-Source
PM04A	A08	Adenosine 2'-Monophosphate	P-Source
PM04A	A09	Adenosine 3'-Monophosphate	P-Source
PM04A	A10	Adenosine 5'-Monophosphate	P-Source
PM04A	A11	Adenosine 2',3'-Cyclic Monophosphate	P-Source
PM04A	A12	Adenosine 3',5'-Cyclic Monophosphate	P-Source
PM04A	B01	Thiophosphate	P-Source
PM04A	B02	Dithiophosphate	P-Source

PM04A	B03	D,L-a-Glycerol Phosphate	P-Source
PM04A	B04	b-Glycerol Phosphate	P-Source
PM04A	B05	Carbamyl Phosphate	P-Source
PM04A	B06	D-2-Phospho-Glyceric Acid	P-Source
PM04A	B07	D-3-Phospho-Glyceric Acid	P-Source
PM04A	B08	Guanosine 2'-Monophosphate	P-Source
PM04A	B09	Guanosine 3'-Monophosphate	P-Source
PM04A	B10	Guanosine 5'-Monophosphate	P-Source
PM04A	B11	Guanosine 2',3'-Cyclic Monophosphate	P-Source
PM04A	B12	Guanosine 3',5'-Cyclic Monophosphate	P-Source
PM04A	C01	Phosphoenol Pyruvate	P-Source
PM04A	C02	Phospho-Glycolic Acid	P-Source
PM04A	C03	D-Glucose-1-Phosphate	P-Source
PM04A	C04	D-Glucose-6-Phosphate	P-Source
PM04A	C05	2-Deoxy-D-Glucose 6-Phosphate	P-Source
PM04A	C06	D-Glucosamine-6-Phosphate	P-Source
PM04A	C07	6-Phospho-Gluconic Acid	P-Source
PM04A	C08	Cytidine 2'-Monophosphate	P-Source
PM04A	C09	Cytidine 3'-Monophosphate	P-Source
PM04A	C10	Cytidine 5'-Monophosphate	P-Source
PM04A	C11	Cytidine 2',3'-Cyclic Monophosphate	P-Source
PM04A	C12	Cytidine 3',5'-Cyclic Monophosphate	P-Source
PM04A	D01	D-Mannose-1-Phosphate	P-Source
PM04A	D02	D-Mannose-6-Phosphate	P-Source
PM04A	D03	Cysteamine-S-Phosphate	P-Source
PM04A	D04	Phospho-L-Arginine	P-Source
PM04A	D05	O-Phospho-D-Serine	P-Source
PM04A	D06	O-Phospho-L-Serine	P-Source
PM04A	D07	O-Phospho-L-Threonine	P-Source
PM04A	D08	Uridine 2'-Monophosphate	P-Source
PM04A	D09	Uridine 3'-Monophosphate	P-Source
PM04A	D10	Uridine 5'-Monophosphate	P-Source
PM04A	D11	Uridine 2',3'-Cyclic Monophosphate	P-Source
PM04A	D12	Uridine 3',5'-Cyclic Monophosphate	P-Source
PM04A	E01	O-Phospho-D-Tyrosine	P-Source
PM04A	E02	O-Phospho-L-Tyrosine	P-Source
PM04A	E03	Phosphocreatine	P-Source
PM04A	E04	Phosphoryl Choline	P-Source
PM04A	E05	O-Phosphoryl-Ethanolamine	P-Source
PM04A	E06	Phosphono Acetic Acid	P-Source
PM04A	E07	2-Aminoethyl Phosphonic Acid	P-Source
PM04A	E08	Methylene Diphosphonic Acid	P-Source
PM04A	E09	Thymidine 3'-Monophosphate	P-Source
PM04A	E10	Thymidine 5'-Monophosphate	P-Source
PM04A	E11	Inositol Hexaphosphate	P-Source
PM04A	E12	Thymidine 3',5'-Cyclic Monophosphate	P-Source
PM04A	F01	Negative Control	S-Source
PM04A	F02	Sulfate	S-Source
PM04A	F03	Thiosulfate	S-Source
PM04A	F04	Tetrathionate	S-Source
PM04A	F05	Thiophosphate	S-Source
PM04A	F06	Dithiophosphate	S-Source

PM04A	F07	L-Cysteine	S-Source
PM04A	F08	D-Cysteine	S-Source
PM04A	F09	Cys-Gly	S-Source
PM04A	F10	L-Cysteic Acid	S-Source
PM04A	F11	Cysteamine	S-Source
PM04A	F12	L-Cysteine Sulfinic Acid	S-Source
PM04A	G01	N-Acetyl-L-Cysteine	S-Source
PM04A	G02	S-Methyl-L-Cysteine	S-Source
PM04A	G03	Cystathionine	S-Source
PM04A	G04	Lanthionine	S-Source
PM04A	G05	Glutathione	S-Source
PM04A	G06	D,L-Ethionine	S-Source
PM04A	G07	L-Methionine	S-Source
PM04A	G08	D-Methionine	S-Source
PM04A	G09	Gly-Met	S-Source
PM04A	G10	N-Acetyl-D,L-Methionine	S-Source
PM04A	G11	L-Methionine Sulfoxide	S-Source
PM04A	G12	L-Methionine Sulfone	S-Source
PM04A	H01	L-Djenkolic Acid	S-Source
PM04A	H02	Thiourea	S-Source
PM04A	H03	1-Thio-b-D-Glucose	S-Source
PM04A	H04	D,L-Lipoamide	S-Source
PM04A	H05	Taurocholic Acid	S-Source
PM04A	H06	Taurine	S-Source
PM04A	H07	Hypotaurine	S-Source
PM04A	H08	p-Aminobenzene Sulfonic Acid	S-Source
PM04A	H09	Butane Sulfonic Acid	S-Source
PM04A	H10	2-Hydroxyethane Sulfonic Acid	S-Source
PM04A	H11	Methane Sulfonic Acid	S-Source
PM04A	H12	Tetramethylene Sulfone	S-Source

Chapter 10. Appendix C - The draft of the paper (Rattleff, et al., 2013b)

Included is the current the draft of the paper

(Rattleff, Nielsen, Rokem, & Lantz, 2013b)

In its current state

Metabolic flux profiling of sixteen *Actinobacteria* species

Stig Rattleff¹, Jens Nielsen², and J. Stefan Rokem³, Anna Eliasson Lantz^{1,*}

¹Center for Microbial Biotechnology, Building 223, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

²Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

³Permanent address: Department of Molecular Genetics & Biotechnology, The Hebrew University of Jerusalem, IL-91120 Jerusalem, Israel

Running title: Central carbon metabolism in *Actinobacteria*

Keywords: central carbon metabolism, metabolic network analysis, isotope labeling, ED, *Actinobacteria*

*Corresponding author:

Anna Eliasson Lantz

Address see above

Telephone +45 4525 2680

Facsimile +45 4588 4148

E-mail ael@bio.dtu.dk

Abstract

The phylum *Actinobacteria* includes many bacteria of industrial importance for production of primary and secondary metabolites. Both the formation of primary and secondary metabolites depends on precursors and cofactors that are provided by the central carbon metabolism of microorganisms. There are alternative pathways for catabolism of carbon either via the Embden Meyerhof Parnas (EMP) and pentose phosphate (PP) pathway or through the Entner Doudoroff (ED) pathway. The EMP pathway is energetically more favorable and has therefore been presumed to be the dominating route for carbon metabolism in bacteria producing secondary metabolites. However, the primary metabolism is poorly studied for most *Actinobacteria* species, as focus traditionally has been on secondary metabolism. With the aim to gain more knowledge about the diversity of central carbon metabolism within the phylum *Actinobacteria*, 16 different species, representing four families were investigated. The study included 11 species belonging to the family Streptomycetaceae, three belonging to Pseudonocardiaceae, one belonging to Micromonosporaceae and one species belonging to the family Corynebacteriaceae. ^{13}C -labeling methodology was applied to determine active pathways and to estimate fluxes through the central carbon metabolism. Four of the 16 species were found to have an active ED pathway, with three of these species belonging to the family Pseudonocardiaceae for which there has not earlier been reported presence of an active ED pathway, and one specie belonging to the family Streptomycetaceae. Clustering of the strains based on activities in the central carbon metabolism does not correlate with their phylogeny. Hence, the results indicate that the pathway in use for central carbon metabolism is more connected to the environmental niche the different species have evolved in rather than their phylogenetic group.

Introduction

The phylum Actinobacteria contains many species with high industrial relevance as it includes many antibiotic producers. Many of the known strains have been discovered in the search for new antibiotics and other bioactive compounds and more than half of the known microbial antibiotics are produced by *Streptomyces* spp. (Kieser et al 2000). Actinobacteria are also used for the production of enzymes, herbicides and amino acids. Consequently, there has been a great interest in studying this group of bacteria. Focus has, however, been on biosynthetic pathways leading to secondary metabolite formation and their regulation, and the central carbon metabolism is still poorly understood for many Actinobacteria [Hodgson, 2000 #72], Janacek 1997.

There are three different pathways by which glucose can be converted to pyruvate: the Embden-Meyerhof-Parnas (EMP) pathway, the Entner-Doudoroff (ED) pathway and the pentose phosphate (PP) pathway. In streptomycetes, glucose has been considered to be metabolized through the EMP pathway and the PP pathway, the former normally being the dominating one [Hodgson, 2000 #72](Cochrane 1961). ED pathway activity is regarded as very rare within the phylum *Actinobacteria*. Until recently the presence of ED pathway activity had only been discovered in one single mycobacterium species, *Mycobacterium smegmatis* (Bai et al 1976), but in the recent years our perception of glucose metabolism in *Actinobacteria* has been revised with reports on ED pathway activity in *Nonomuraea* sp. ATCC 39727 [Gunnarsson, 2004 #69] and in *Str. tenebrarius*. [Borodina, 2005 #67].

The objective of the present study was to gain more knowledge about the diversity of central carbon metabolism within the phylum *Actinobacteria*. In particular we wanted to address the question how wide spread ED pathway activity is in *Actinobacteria* and whether there is a correlation between phylogeny and activities of the central carbon pathways Fuhrer et al 2005, Selig et al 1997, Rominus and Morgan 2003. Sixteen strains representing four families, namely

Streptomycetaceae, Micromonosporaceae, Pseudonocardiaceae and Corynebacteriaceae, were collected from various sources and strain collections. The strains were grown in minimal medium with [1-¹³C] glucose as carbon source. The ¹³C-labeling patterns of the proteinogenic amino acids were determined by GC-MS analysis. The resulting isotope patterns reflect the activity of intracellular pathways and the network topology of active reactions can be identified. There are several recent examples where the use of isotope balancing in combination with metabolite balancing has been shown to be useful for identification of novel or unexpected metabolic pathways (Fischer 2003) [Gunnarsson, 2004 #69] + wiechert?). In the present study the labeling patterns of alanine and valine were of particular interest as they are derived from pyruvate and therefore can be used to distinguish whether glucose is metabolized through the ED pathway, the EMP pathway or the PP pathway.

In addition to obtain fundamental information about the inheritance of central carbon pathways in *Actinobacteria* we wanted to gain a better understanding of the central carbon metabolism in general. This would provide knowledge that can be used for strain improvement in antibiotic production processes as the central carbon metabolism is tightly linked to secondary metabolism through supply of building blocks/precursors and co-factors.

Materials and Methods

Strain and propagation. The following strains were used: *Streptomyces alboniger* ATCC 12462 (received from The Hebrew University of Jerusalem), *Str. avermitilis* DSM 46492 (ordered from DSMZ), *Str. canescens* ATCC 15731 (received from The Hebrew University of Jerusalem), *Str. coelicolor* M145 (received from John Innes Centre, Norwich, UK), *Str. mediterraneus* ATCC 13278 (received from The Hebrew University of Jerusalem, Israel), *Str. refuineus* NRRL 3143 (received from The Hebrew University of Jerusalem, Israel), *Str. roseochromogenes* ATCC 13400 (received from The Hebrew University of Jerusalem, Israel), *Str. tenebrarius* ATCC 17920 (ordered from ATCC), *Str. tsusumacensis* ATCC 15141 (received from The Hebrew University of Jerusalem, Israel), *Str. PRL 1642* ATCC 23836 (received from The Hebrew University of Jerusalem, Israel), *Actinoplanes teichomyceticus* ATCC 31121 (ordered from ATCC), *Amycolatopsis balhimycina* DSM 5908 (received from University of Tübingen, Germany), *Am. orientalis* ATCC 19795 (ordered from ATCC), *Saccharopolyspora erythraea* DSM 40517 (ordered from DSMZ), *Corynebacterium glutamicum* DSM 20300 (ordered from DSMZ), and *Streptomyces lividans* TK24 (received from Katholieke Universiteit Leuven, Belgium). The strains were maintained on YESP agar plates (glucose 15 g l⁻¹, glycerol 15 g l⁻¹, yeast extract 5 g l⁻¹, soya peptone 15 g l⁻¹, NaCl 3 g l⁻¹, agar 20 g l⁻¹, pH=7.0). For longer storage, stocks (as spore suspensions or mycelia) were stored at -80 °C in 20% glycerol. For seed culture preparations, a 500 mL baffled Erlenmeyer flask containing 50 mL YESP medium was inoculated with one loop cells from a plate and incubated with shaking (150 rpm) at 28°C for 24-48 h until a OD_{600nm} of 4-6 was reached.

Cultivation conditions. The cultivations with labeled substrate were carried out in 500 mL Erlenmeyer flasks containing 100 mL of a minimal medium. The medium contained 9.9 g l⁻¹ (NH₄)₂SO₄, 1.63 g l⁻¹ KH₂PO₄, 1.5 g l⁻¹ MgSO₄ · 7H₂O, 1 g l⁻¹ NaCl, 20 mg l⁻¹ FeSO₄ · 7H₂O, 20

mg l⁻¹ ZnSO₄ · 7H₂O, 10 mg l⁻¹ MnSO₄ · H₂O, 10 mg l⁻¹ CaCl₂ · 2H₂O, 5 mg l⁻¹ Na₃-citrate · 2H₂O, 50 mM MES and 20 g l⁻¹ glucose, with an initial pH of 7.0. Half of the glucose was [1-¹³C] glucose (Omicron Biochemicals, South Bend, IN). The flasks were inoculated to a start OD_{600nm} of 0.3 from the seed cultures and incubated at 28°C and 150 rpm. Each strain was cultivated twice.

Cell mass determination. Throughout the cultivations, cell mass concentration was followed by measurement of OD_{600nm}. For the end samples dry cell mass was determined gravimetrically. Volumes between 5 and 10 ml, depending on the concentration of cell mass, were filtered through pre-dried nitrocellulose filters with pore size 0.45 µm (Supor-450; Pall Corporation, Ann Arbor, MI). The filter cakes were washed with distilled water, twice the sample volume, and dried in a microwave oven for 20 min at 150W. The mass gain was measured after the filter was cooled down in a desiccator for 15-30 minutes.

Amino acid derivatization and GC-MS analysis. For analysis of proteinogenic amino acids, cell mass was harvested by centrifugation (5000g, 10 minutes), washed twice with distilled water and stored at -20°C until further analysis. Samples, in duplicates, were withdrawn from the cultivations at two time-points, corresponding to mid exponential and late exponential growth phase. The day before analysis the cell mass was dried in aliquots 16-20 h at 105°C. For hydrolysis of total cell protein, approx. 5 mg of dry cell mass was resuspended in 800 µl 6M HCl and hydrolyzed for 16-24h at 105°C. Next the hydrolyzate was divided into 200 µl aliquots, dried at 105°C and used for amino acid derivatization. Amino acid derivatization and GC-MS analysis were performed as described previously (Christensen and Nielsen, 1999, 2000).

Several ion clusters can be seen for each amino acid in the MS spectra, due to fragmentation of the derivatized amino acids during the ionization process. The labeling of each fragment was corrected for the natural labeling of the derivatization agent and for natural labeling of oxygen, hydrogen and

nitrogen content of the amino acids and the labeling patterns of the amino acid derived parts of the fragments were described in terms of summed fractional labeling, SFL (Christensen and Nielsen, 1999, 2000), according to the formula:

$$\text{SFL (\%)} = \frac{0 \cdot I_0 + 1 \cdot I_1 + \dots + n \cdot I_n}{I_0 + I_1 + \dots + I_n} \times 100$$

where I_0 is the intensity of the peak resulting from the unlabeled mass isotopomer, I_1 is the intensity of the peak arising from the mass isotopomers with one ^{13}C atom, and I_n is the intensity of the peak corresponding to the mass isotopomer where all carbon atoms are ^{13}C . As the natural occurrence of ^{13}C is 1.1%, the SFL of a naturally labeled fragment with n carbon atoms will be $n \cdot 1.1\%$.

Data analysis. The software Unscrambler version 7.6 (CAMO, Norway) was used for principal component analysis (PCA) modeling. SFL values for all amino acid fragments except those of threonine and serine were used as predictors (X matrix). All X-variables were centered and weighted to unit variance. Segmented cross validation was applied with samples from the same strain forming one block. PCA decomposes data into a structure part and a noise part in order to provide an interpretable overview of the main information of multivariate data (Wold et al., 1987). The algorithm reduces the number of variables and the information is projected onto a smaller number of significant variables, the so-called principal components (PCs). The PCs are linear combinations of the original variables and are selected so that the first PC covers as much of the variation in the data as possible. The second PC is orthogonal to the first and covers as much of the remaining variation as possible, and so forth.

For hierarchical clustering the software MATLAB Version 7.0.4 (The MathWorks, Inc., Natick, MA, USA) was applied. Group average linkage (also called UPGMA) was used as agglomerative method and Euclidean distance was used as similarity measure between samples. SFL values were normalized before analysis. Hierarchical clustering works by iteratively joining the two closest

clusters starting from singleton clusters (agglomerative methods) or iteratively partitioning clusters starting with the complete set (divisive methods) (Guenoche 1991). With the linkage method used in the present work, at each stage, two clusters are merged such that the newly formed cluster, on average, will have minimum pairwise distances between its points. Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram which illustrates the fusions or divisions made at each successive stage of analysis.

RESULTS

In order to estimate the activity of different pathways in the central carbon metabolism in *Actinobacteria*, ^{13}C -labeling experiments were performed and the positional labeling patterns of central metabolites determined for sixteen different strains. The strains were selected from different groups of the phylum *Actinobacteria*, however, the majority were streptomycetes, as an active ED pathway recently was detected in a species belonging to this genus, namely *Streptomyces tenebrarius* [Borodina, 2005 #67]. Batch cultivations in minimal media with glucose as the sole carbon source was carried out as described above, with half of the glucose labeled in the first position. Cells were harvested at two time points in the exponential growth phase for each strain and the SFLs of the proteinogenic amino acid fragments were determined by GC-MS analysis (Table 1).

EMP and ED. Glucose can hypothetically be catabolized to triose through the EMP, ED or PP pathways. The activity of each metabolic pathway can be estimated from the positional labeling of pyruvate. The oxidative part of PP and ED both result in unlabeled pyruvate-C3, whereas if glucose

is metabolized through EMP, 50% of the pyruvate will be labeled in position 3 when 100% [1]-¹³C-glucose is used as substrate (Figure PYRlabel). The natural labeling of pyruvate-C1 is 1.1%. Strains having a significant higher labeling in this position of pyruvate have an active ED pathway as conversion of glucose to pyruvate through the ED pathway results in, that the labeling of [1]-¹³C-glucose remains in the carbon in the first position. By combining the SFLs of alanine 158, alanine 99/116 and valine 143, the fractional labeling for each C of pyruvate can be calculated. High labeling in pyruvate-C1, and consequently presence of ED pathway were observed for four strains out of the sixteen investigated; *Sacc. erythrea*, *Am. orientalis*, *Str. tenebrarius* and *Am. balhimycina* (Table 1). *Sacc. erythrea* and *Am. orientalis* almost exclusively used ED, whereas a combination of ED and EMP metabolism was seen for the other two strains, *Str. tenebrarius* and *Am. balhimycina*.

PP pathway. Amino acids originating from precursor metabolites in the lower part of EMP pathway, such as glycerol 3-phosphate (G3P) and pyruvate, have lower labeling for strains having a high PP activity as the carbon 13 from the C1 of glucose is lost as carbon dioxide in the oxidative part of PP. SFL of alanine 158, corresponding to pyruvate C1-3, can be used to estimate the PP activity, the lower the SFL the higher the pentose phosphate pathway (PPP) flux. The SFL of the glucose fed to the cells in the present work was 55.55%, and the labeling of pyruvate C1-3 will then be 27.75% if PPP is inactive. The strains could be divided into three groups based on PPP activity. *Str. sp* PRL1642, *Str. mediocidicus*, *Str. roseochromogenes* and *Str. tsusumacensis* had SFLs of 27-28% indicating very low PP activity under the investigated conditions. *Sacc. erythrea*, *Str. canescus*, *Str. avermitilis*, *Str. refuineus*, *Str. coelicolor*, *Str. lividans*, *Am. balhimycina*, and *Act. teichomyceticus* had a low to moderate PP activity, whereas, high PP activity was observed for *Str. alboniger*, *Str. tenebrarius*, *C. glutamicum* and *Am. orientalis*.

TCA cycle and anaplerotic routes. C2 of aspartate, derived from oxaloacetate, is formed from C3 and C4 of α -ketoglutarate. C3 of pyruvate (comes from glucose-C1 if the EMP pathway is active) corresponds to C4 in α -ketoglutarate. A high SFL of aspartate¹¹⁵, which represents labeling of C2 in oxaloacetate, therefore corresponds to a high TCA flux. For the strains having an active ED pathway the labeling in aspartate 115 was low as expected, as pyruvate-C1 is lost as CO₂ resulting in a low labeling of AcCoA. Thus, one cannot conclude anything about the TCA flux using the SFL of aspartate¹¹⁵ for strains having ED pathway activity.

Investigation of the labeling in pyruvate and TCA-intermediates also provides information about the presence of anaplerotic routes. If the labeling in TCA-intermediates is solely from pyruvate, the labeling per C of pyruvate should be similar to that of oxaloacetate and α -ketoglutarate. This was the case for the non-ED strains, and consequently, there was no (or low) flux through anaplerotic routes at the investigated conditions (Table 1). For the two strains with high ED activity, *Am. orientalis* and *Sacc. erythrea*, a higher labeling in TCA-intermediates than in pyruvate was observed. The labeling per carbon of AcCoA entering TCA was 2.8 and 2.3, whereas the TCA-intermediates had labelings of 4.0-4.1 and 4.0-4.3 per carbon, for *Am. orientalis* and *Sacc. erythrea*, respectively, indicating that carbon must enter the TCA through another route. PEP carboxylase and pyruvate carboxylase are both anaplerotic enzymes that have been found in *Actinobacteria* and have been thoroughly studied in *Corynebacterium glutamicum* (Shirai 2007 AND Peters-Wendish 2001 AND Mukhopadhyay 2000). As the labeling per carbon in PEP was 1.6 and 1.7, respectively, the presence of PEP carboxylase would not result in increased labeling in TCA-intermediates. However, through the action of pyruvate carboxylase, catalysing the carboxylation of pyruvate to oxaloacetate, the labeling from pyruvate-C1 (which is high for strains with ED activity) will be transferred to oxaloacetate.

Malic enzyme. Malic enzyme converts malate to pyruvate coupled with the production of NAD(P)H and has been found in *C. glutamicum* (Gourdon 2000), in *Mycobacterium*, (probable NAD enzyme found in *Mycobacterium bovis* and *smegmatis* BLAST), and in *Str. coelicolor* (Rodriguez 2012). From the SFL of phenylalanine¹⁴³ it could be seen that the labeling per carbon in PEP was close to the natural labeling for all strains (1.5-1.9). On the other hand, the labeling in pyruvate-C1,C2 (valine¹⁴³) for many of the strains was higher than the natural labeling. A higher labeling in pyruvate-C1 can be the result of an active ED pathway as discussed earlier, so instead the labeling in pyruvate-C2, calculated from alanine and valine fragments was assessed. An increased labeling (4.4-4.8) was observed for *Str. sp* PRL1642, *Str. mediodidicus*, *Str. roseochromogenes* and *Str. tsusumacensis*, indicating activity of malic enzyme in those strains.

Clustering of strains. In order to examine data structure and evaluate whether the grouping of strains based on central carbon metabolism activity follows phylogeny or not, PCA and hierarchical clustering were applied on the obtained SFLs for the different strains. PCA is a data reduction method, which from multivariate data defines a smaller variable set, so called principal components (PCs), that retains the amount of information present in the original data set (Wold 1987). The two first principal components explained 89% of the variation and the strains formed 4 clusters (Figure PCA). Three of the strains that have an active ED pathway, *Sac. erythrae*, *Am. orientalis* and *Str. tenebrarius*, formed a cluster, whereas the fourth strain with ED, *Am. balhimycina*, did not cluster with these three strains, but with cluster three, which consisted of *C. glutamicum*, *Str. canescus* and *Str. alboniger*. Cluster two comprised *Act. teichomycetius*, *Str. lividans*, *Str. coelicolor*, *Str. avermitlitis*. *Str. refuineus*; *Str. mediodidicus*, *Str. PRL 1642*, *Str. roseochromogenes* and *Str. tsusumacensis* formed the last cluster. Comparing the score plot with the loading plot (data not shown) it could be seen that the direction of PC1 corresponded to the

labeling in pyruvate-C3. To verify the influence of an active ED pathway on the grouping of strains, the analysis was redone without the SFLs of valine¹⁴³ and ¹⁸⁶, glycine¹⁷⁵, alanine¹⁵⁸ and phenylalanine¹⁴³ and ¹⁹². The same clustering as before could be observed, with the only difference that *Am. balhimycina* did not form a separate group any more in the score plot, but clustered with *C. glutamicum*, *Str. canescus* and *Str. alboniger* (data not shown). This indicates that clustering was based on total central carbon activity and not only on EMP and ED activities.

The second unsupervised method applied, hierarchical clustering, uses dissimilarities or distances between the objects when forming clusters. Hierarchical Clustering is subdivided into bottom-up or agglomerative methods, which start with each object in a separate cluster and by series of fusions of the *n* objects combines them into groups, and top-down or divisive methods, which separate *n* objects successively into finer groupings. We have used an agglomerative method, group average, where the distance between two clusters is calculated as the average distance between all pairs of objects in the two different clusters. The dendrogram plot of the clustering results suggested 4 clusters, the clusters being the same as for PCA (figure dendrogram). If the cut off was chosen to gain 5 clusters, *Am. balhimycina* formed a cluster by its own, and with 4 clusters it grouped with *C. glutamicum*, *Str. canescus* and *Str. alboniger*, as for the PCA without SFLs for pyruvate derived amino acid fragments.

The specific growth rate was determined for the different strains (data not shown). It could be argued that the effect of the growth rate would dominate the clustering results, though it was clear from the clustering that this was not the case, as there was a considerable variance of growth rates within each of the obtained clusters (data not shown).

DISCUSSION

We have applied ^{13}C -labeling methodology for mapping of central carbon metabolism in sixteen different *Actinobacteria* strains. The use of labeled glucose for analysis of network topology is an attractive method since it provides integrative information about the activity of different pathways and thereby about *in vivo* fluxes in the cells. Furthermore, in contrast to other methods applied in metabolic engineering, e.g. transcriptome and proteome analysis, ^{13}C -labeling can be applied on micro-organisms for which there is only little knowledge available.

PCA as well as hierarchical clustering based on the SFL data directly resulted in four to five clusters. To explore which common traits in physiology that made the strains cluster together, the level of activity of the ED, EMP, PP and TCA pathways were estimated for all the investigated strains based on SFL values as discussed earlier (Table 2). Characteristics of cluster one (*Sac. erythrae*, *Am. orientalis* and *Str. tenebrarius*) was in particular presence of an active ED pathway, furthermore, the strains had low EMP activity. The reason why *Am. balhimycina* clustered with cluster three might be due to the fact that this strain also has a higher EMP activity than the other strains having ED pathway activity. The strains in cluster three (*C. glutamicum*, *Str. canescus* and *Str. alboniger*) all displayed medium EMP activity and low TCA activity. Typical traits for strains in cluster two (*Act. teichomycetius*, *Str. lividana*, *Str. coelicolor* and *Str. avermitilis*. *Str. refuineus*) were medium EMP activity, low PPP activity and medium/high TCA activity. Finally, the strains in cluster four (*Str. mediocidicus*, *Str. PRL 1642*, *Str. roseochromogenes* and *Str. tsusumacensis*) were those that showed the highest EMP and the lowest PPP activity among the strains tested.

We identified an active ED pathway in three strains for which the presence of this pathway has not been reported before; two glycopeptide producing strains (*Am. balhimycina* and *Am. orientalis*) and the erythromycin producer (*Sac. erythrae*). While this work was on-going, the genomic sequence of *Sac. erythrae* became available; in this sequence the two genes have been found and annotated (Olinyinyk 2007, and Licona -Cassani 2010), though so far no functional proof has been published. The presence of two genes with high similarity (Blastn alignment, Altschul 1997) to *eda* and *edd* of *Nonomuraea* (GenBank accession number AJ6055554 and AJ6055555, respectively) indicates the presence of this pathway in a number of *Actinobacteria*, including the two Streptomycetes, *Str. hygroscopicus subsp. jinggangensis* and *Str. cattleya*, as well as *Sac. Polyspora*, which has been published, *Am. Mediterranei* which is a close relative of the two published *Amycolatopsis spp.*, and *M. smegmatis* (see supplementary table 1). In the ED pathway glucose 6P is converted to 6-phosphogluconate instead of to fructose 6P as in the EMP pathway. 6-phosphogluconate then goes to 2-keto-3-deoxy-6-phosphogluconate (KDPG) and subsequently KDGP is split into two C3 molecules. In summary from one mole glucose, 1 mole NADPH, 1 mole NADH and 1 mole ATP are produced when the ED pathway is used, whereas the EMP pathway yields 2 moles NADH and 2 moles ATP per mole glucose. Since the ED pathway is energetically less favorable, one may wonder what the rationale behind the usage of this pathway is? In *E.coli*, the ED pathway is believed to be central to utilization of gluconate and other sugar acids [Conway, 1992 #40 / Peekhaus, 1998 #37 / Murray, 2005 #42]. Moreover, Eda synthesis has been seen to be induced by phosphate limitation and carbon starvation (Murray 2005). One could also speculate that the higher NADPH production during operation of the ED pathway compared to when the EMP pathway is employed would be beneficial for antibiotic production where NADPH often functions as an important cofactor.

Among the strains investigated in this study *Str. tenebrarius* was the only streptomycete found to express an active ED pathway. One could start to wonder if the tobramycin producer designated *Str. tenebrarius* really is a streptomycete. Streptomycetes can be distinguished from other actinomycetes by their cell wall type (Type I) characterized by the presence of ll-diaminopimelic acid and glycine (Lechevalier & Lechevalier, 1970). The cell wall of *Str. tenebrarius* has been found to contain meso-diaminopimelic acid (Tamura 2008), which indicates that it indeed is not a true streptomycete.

The ED pathway is believed to evolutionary pre-date the EMP (Melendez-Hevia 1997; Romano and Conway 1996). We had expected to see a correlation between phylogeny and activity of central carbon metabolism. Surprisingly, no such trend could be detected. Even though the three strains discovered to have ED activity in the present study all belong to the Pseudonocardiaceae family; earlier known species expressing ED activity were widespread among different families: *Str. tenebrarius* of the family Streptomycetaceae, *Nonomuraea* sp. of the family Streptosporangiaceae and *Mycobacterium smegmatis* of the family Mycobacteriaceae. Furthermore, streptomycetes could be found in all clusters and neither were the closely related *A. balhimycina* and *A. orientalis* adjacent. The present findings indicate that activity in the central carbon metabolism is more connected to environmental niche than phylogenetic group, which could be explained by the considerable horizontal gene transfer within bacteria. A similar pattern has been observed for secondary metabolite genes. Metsä-Ketelä et al (2002) investigated the molecular evolution of aromatic polyketides in *Actinobacteria*. A phylogenetic analysis showed discrepancy between a polyketides sequences based clustering and a 16S based one, and they

concluded that horizontal transfer of polyketides genes was probably the reason for the disagreement.

Acknowledgement

References:

Tables and Figures

TABLE 1. Measured summed fractional labellings for samples taken in the exponential growth phase of the fifteen strains.

Compound analyzed	MW fragment	Biosynthetic precursor(s) ^a	Carbon atoms ^b	Sum of fractional labelings (%) of C atoms ^c															
				<i>Sacc. erythraea</i>	<i>Am. orientalis</i>	<i>Str. tenebrarius</i>	<i>Am. balthimycina</i>	<i>Str. canescens</i>	<i>Str. alboniger</i>	<i>C. glutamicum</i>	<i>Str. lividans</i>	<i>Str. coelicolor</i>	<i>Act. telchomyces</i>	<i>Str. refulgens</i>	<i>Str. avermitidis</i>	<i>Str. PRL1642</i>	<i>Str. mediodidicus</i>	<i>Str. roseochromoge</i>	<i>Str. tsusumace</i>
Glycine	175	3-PG	1, 2	3.7	3.6	6.6	4.2	5.9	3.7	3.0	4.4	4.5	5.3	3.7	3.3	5.1	4.6	3.8	4.8
Glycine	144	3-PG	1, 2	3.5	3.6	6.8	3.0	4.9	3.1	2.7	52.8	2.7	3.5	4.2	3.6	3.9	4.1	4.2	3.9
Glycine	85	3-PG	2	1.8	1.3	4.6	2.5	4.1	1.5	1.9	0.8	2.5	3.4	2.5	1.5	3.5	2.3	1.9	2.9
Serine	175	3-PG	1, 2	ND ^d	4.4	5.3	4.3	5.3	3.9	7.4	2.5	5.4	5.4	4.8	10.0	5.4	5.3	10.4	5.5
Serine	132	3-PG	2, 3	ND	6.8	8.1	15.7	14.0	11.5	15.2	14.7	17.6	16.5	17.8	18.0	23.1	22.4	24.1	23.3
Phenylalanine	143	PEP	1, 2	3.4	3.3	3.4	3.6	3.1	3.0	3.6	5.4	3.3	3.8	3.6	3.4	3.9	3.8	3.8	3.8
Phenylalanine	192	PEP + B4P	2, 2, 3, 3 + 1, 2, 3, 4	15.8	24.4	23.4	42.3	16.7	32.7	31.5	40.8	41.6	45.1	49.2	47.9	60.3	63.2	62.0	61.2
Alanine	116	Pyr	2, 3	4.2	5.8	8.3	15.5	17.9	13.0	14.5	16.6	18.7	18.8	19.0	18.5	25.4	26.3	25.2	25.5
Alanine	99	Pyr	2, 3	5.1	5.5	8.5	15.0	18.5	12.8	14.4	15.8	18.1	18.7	19.1	18.6	24.9	26.0	25.4	25.2
Alanine	158	Pyr	1, 2, 3	19.6	17.6	15.9	22.4	20.0	14.0	16.3	20.3	20.7	23.8	20.6	20.1	27.8	28.7	27.7	28.0
Valine	144	Pyr	2, 2, 3, 3	8.0	10.9	14.9	27.4	26.4	24.5	29.3	31.2	33.9	32.4	35.7	36.9	48.5	50.3	49.0	48.4
Valine	143	Pyr	1, 2	17.0	13.8	8.7	9.9	3.9	3.7	3.8	2.6	5.2	4.5	4.1	4.5	7.1	7.0	7.3	7.0
Valine	127	Pyr	2, 2, 3, 3	8.8	10.8	14.8	26.6	27.1	24.1	29.5	31.2	33.1	32.2	35.6	36.7	47.4	50.3	49.3	48.3
Valine	186	Pyr	1, 2, 2, 3, 3	24.2	22.7	21.3	33.4	28.0	25.3	30.0	34.9	35.2	34.2	37.2	38.1	50.1	52.4	51.7	51.1
Leucine	158	Pyr + AcCoA	2, 2, 3, 3 + 2	10.3	15.4	17.9	35.2	22.5	34.4	34.2	45.3	45.6	44.8	50.6	52.7	67.2	70.9	68.8	67.5
Asparatate	188	OAA	2, 3, 4	10.7	12.1	13.9	24.6	23.7	22.2	25.0	29.0	29.5	34.7	34.6	34.9	39.7	41.4	40.0	41.1
Asparatate	115	OAA	2	2.5	3.0	4.4	6.9	8.6	7.5	6.3	9.9	10.3	13.1	12.9	13.3	14.3	15.2	14.9	14.9
Asparatate	216	OAA	1, 2, 3, 4	17.1	16.5	19.8	31.0	29.4	26.3	29.8	34.2	35.2	42.2	41.4	42.1	47.4	49.9	49.4	48.8
Threonine	175	OAA	1, 2	ND	8.9	7.4	12.4	12.9	12.1	10.9	16.5	14.6	18.1	18.9	19.5	20.4	21.7	21.3	21.2
Threonine	146	OAA	2, 3, 4	ND	12.3	11.5	20.7	25.9	21.7	24.6	16.9	26.0	31.5	32.9	33.5	35.5	38.5	37.9	37.3
Isoleucine	158	OAA + Pyr	2, 3, 4 + 2, 3	11.3	15.3	18.9	35.1	29.8	33.8	34.3	44.8	44.9	46.1	50.7	52.6	65.3	69.2	67.5	66.3
Proline	142	a-KG	2, 3, 4, 5	8.6	12.5	14.8	27.3	24.8	28.6	34.8	37.2	35.6	41.0	43.8	44.3	52.2	55.7	52.5	53.0
Glutamate	143	a-KG	1, 2	13.0	9.6	11.4	17.6	17.9	14.9	19.5	17.1	19.2	22.5	22.2	22.7	26.3	27.5	27.1	26.7
Glutamate	230	a-KG	1, 2, 3, 4, 5	14.0	16.1	23.0	37.7	43.1	33.2	39.3	45.7	44.3	53.2	53.6	52.8	65.3	67.6	66.4	65.9
Lysine	156	a-KG + AcCoA	2, 3, 4, 5 + 2	14.0	16.6	17.1	33.5	25.1	32.3	32.6	43.7	41.8	45.2	48.5	50.9	57.3	61.5	60.1	58.2

^a The biosynthetic precursors indicated are for growth on glucose and ammonia; Abbreviations: 3-PG, 3-phosphoglycerate; Pyr, pyruvate; AcCoA, acetyl-CoA; OAA, oxaloacetate; a-KG, a-ketoglutarate; PEP, phosphoenolpyruvate; B4P, erythrose-4-phosphate; Glc, glucose; Glc 6P, glucose-6-phosphate.

^b Carbon atoms refer to carbon atoms of the precursor metabolite present in the amino acid fragment, denoted by their position in the precursor metabolite.

^c Data are mean values. A naturally labeled fragment with n carbon atoms will be labeled n.n%.

^d ND, not determined

TABLE 2. Estimated activities of central carbon metabolism and growth rates of the investigated strains.

Strain	Family ¹	EMP	ED	PP	TCA ²
<i>Sacc. erythrea</i>	P	(+)	++	+	
<i>Am. orientalis</i>	P	(+)	++	++	
<i>Str. tenebrarius</i>	S	+	+	++	
<i>Am. balhimycina</i>	P	++	+	+	
<i>Str. canescus</i>	S	++	-	+	+
<i>Str. alboniger</i>	S	++	-	++	+
<i>C. glutamicum</i>	C	++	-	++	+
<i>Str. coelicolor</i>	S	++	-	+	+(+)
<i>Str. lividans</i>	S	++	-	+	+
<i>Act. teichomyceticus</i>	M	++	-	+	++
<i>Str. refuineus</i>	S	++	-	+	++
<i>Str. avermitilis</i>	S	++	-	+	++
<i>Str. PRL1642</i>	S	+++	-	(+)	++
<i>Str. mediocidicus</i>	S	+++	-	(+)	++
<i>Str. roseochromogenes</i>	S	+++	-	(+)	++
<i>Str. tsusumacensis</i>	S	+++	-	(+)	++

¹ P: Pseudonocardaceae, S: Streptomyetaceae, C: Corynebacteriaceae, M: Micromonosporaceae

² For the strains having ED activity the TCA activity can not be estimated with the applied method

FIGURES

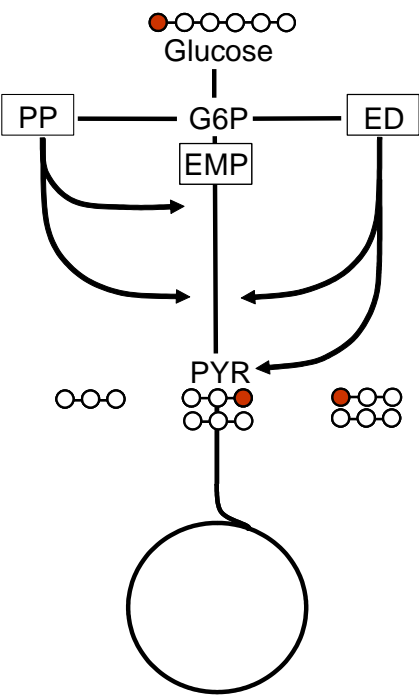


Figure PYRlabel. Illustration of how conversion of glucose into pyruvate through different metabolic pathways result in different positional labeling of pyruvate.

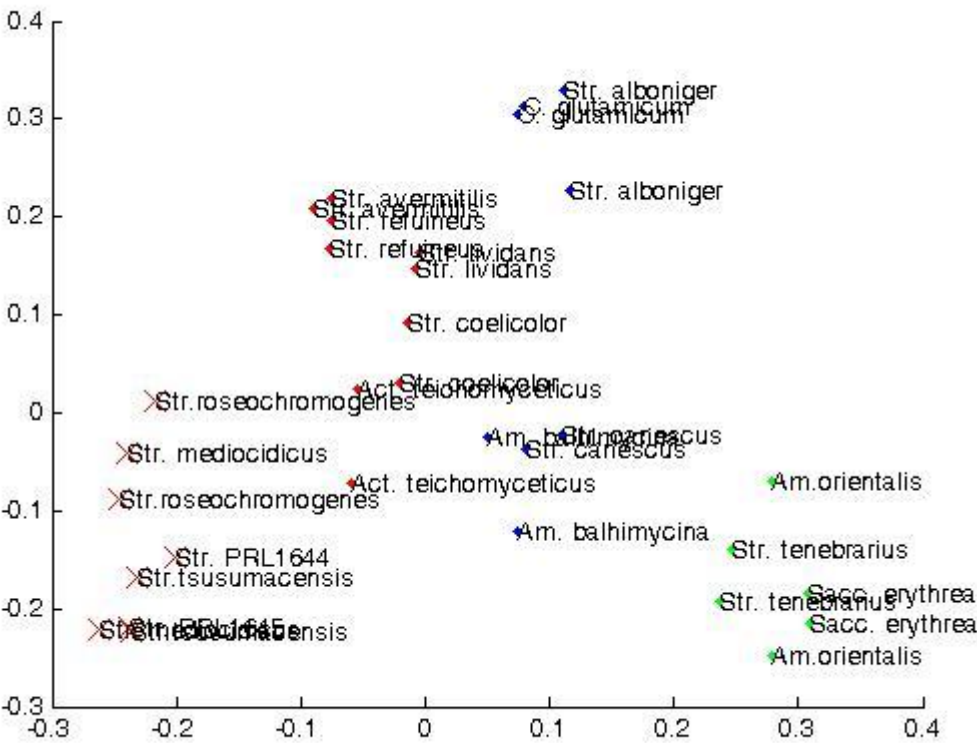


Figure PCA. Biplot of PC1 and PC2 from principal component analysis on SFLs for the different strains. The SFLs for threonine and serine were excluded from the analysis due to high STDdev. The colors and shapes symbolize the group membership, as determined from a K-means clustering algorithm

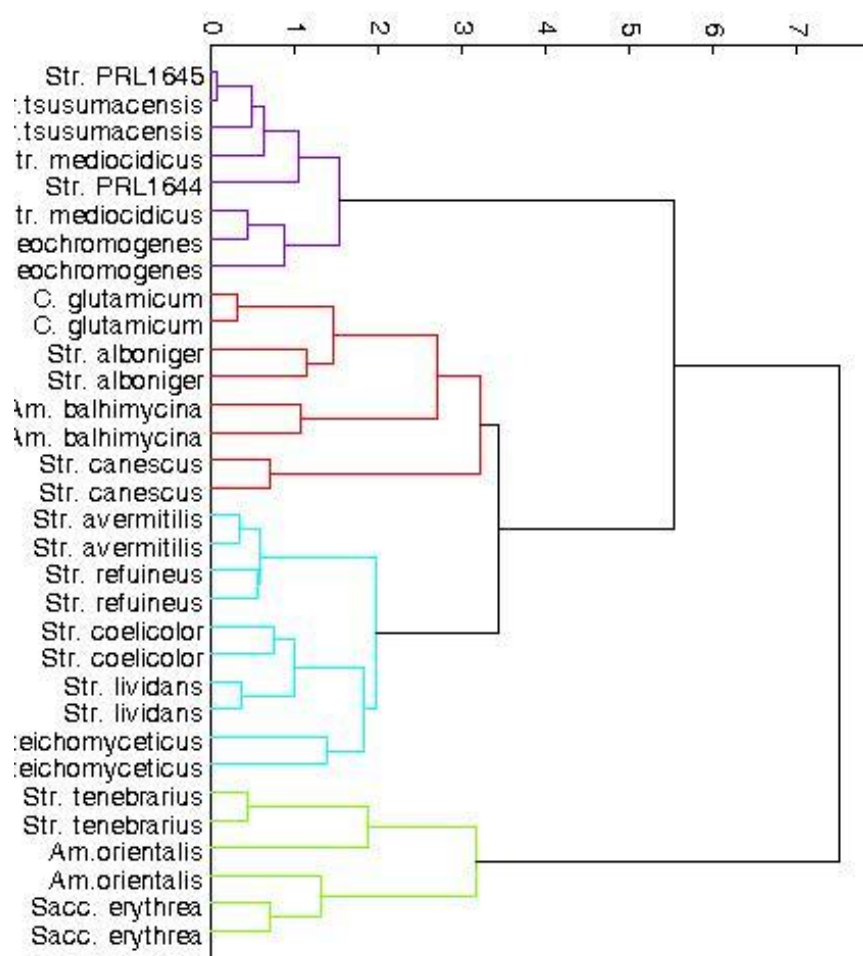


Figure Dendrogram. Dendrogram plot of a hierarchical clustering analysis of SFL data. The SFLs for threonine and serine were excluded from the analysis due to high STDdev.